

# Transcriptome Sequencing and Differential Expression Analysis Reveal Molecular Mechanisms for Seed Starch Accumulation in Chinese Chestnut *Metaxenia*

**Shengxing Li**

Jiangxi Academy of Forestry

**Zhuogong Shi**

Southwest Forestry University

**Zhiheng Zhao** (✉ [zzhhly09@163.com](mailto:zzhhly09@163.com))

Guanxi forestry research institute

**Qiurong Zhu**

Southwest Forestry University

**Liang Tao**

Southwest Forestry University

**Wenhui Liang**

Guanxi forestry reserach instite

---

## Research article

**Keywords:** chestnut, metaxenia, differential gene expression, seed development, starch synthesis

**Posted Date:** November 25th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-80588/v1>

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

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at BMC Genomics on August 13th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07923-5>.

# Abstract

**Background:** Chestnut is an important kind of edible nut rich in starch and protein. The characteristics and nutrient contents of chestnut have been found to show obvious metaxenia effects in previous studies. To improve the understanding of the metaxenia effect on chestnut starch and sucrose metabolism, this study used three varieties of chestnut, 'Yongfeng 1', 'Yong Renzao' and 'Yimeng 1', as male parents to pollinate 'Yongfeng 1', as the female parent, and studied the mechanisms of starch and sucrose metabolism in three starch accumulation stages (70 (S1), 82 (S2), and 94 (S3) days after pollination , DAP) in the chestnut seed kernel.

**Result:** Most carbohydrate metabolism genes were highly expressed in YFF in stage S2 and in YFR and YFM in stage S3. In stage S3, hub genes encoding HSF\_DNA-binding, ACT, Pkinase, and LIM proteins and four transcription factors were highly expressed, with YFF showing the highest expression, followed by YFR and, finally, YFM. In addition, transcriptome analysis of the kernels at 70, 82 and 94 DAP showed that the starch granule-bound starch synthase (EC 2.4.1.242) and ADP-glucose pyrophosphorylase (EC 2.7.7.27) genes were actively expressed at 94 DAF. Chestnut seeds regulate the accumulation of soluble sugars, reducing sugars and starch by controlling glycosyl transferase and hydrolysis activity during development.

**Conclusion:** These studies and resources have important guiding significance for further research on starch and sucrose metabolism and other types of metabolism related to chestnut metaxenia.

## Background

Chinese chestnut (*Castanea mollissima* BL.) has been cultivated for its nuts and timber for over 4000 years [1]. Over 300 cultivars have been selected for nut production[2]. It remains a very important commercial crop in China. As of 2018, Chinese chestnut was being widely grown in Asia, Europe, and America for nut production, accounting for more than 80% of the world's production (adopted from the UN Food and Agriculture Organization, Corporate Statistical Database: <http://www.fao.org/>). Chinese chestnut is a rich source of vitamin C, protein, and a variety of minerals, with particularly high contents of magnesium, phosphorus, and potassium (adopted from USDA FoodData Central: <https://fdc.nal.usda.gov/>). However, Chinese chestnut cultivation continues to face serious problems, such as uneven quality, the production of many hollow bracts and a low yield [3–5]. Chinese chestnut shows typical cross-pollination characteristics and a low self-pollination seed setting rate; therefore, it is very important to separate pollinated trees to improve the chestnut yield[6–7].

The phenomenon in which the endosperm of hybrid modern seeds displays paternal genetic traits is known as xenia[8]. Xenia has applications not only in genetic and physiological research but also in plant breeding and crop production[9]. Xenia includes metaxenia, which is the effect of pollen on fruit size and other fruit characteristics[10–11]. Metaxenia has been described in many species, such as apple[12], maize[13], and blueberry[14]. Previous studies have shown that the traits and composition of chestnut

nut exhibit obvious metaxenia effects[15–17]. In addition, pollination treatments result in significant differences in chestnut ovary nutrition[18]. Therefore, studying the metaxenia effect is very important for the reasonable allocation of chestnut trees used for pollination, for the scientific planning of new chestnut gardens and for improving the quality of chestnut fruit.

Chestnut seed starch accumulation is a very complex process involving many expression-dependent physiological changes and the regulation of a large number of genes and phytohormones. Chinese chestnut flowers in mid-summer, and it takes approximately 100 days for its nuts to fully ripen[19]. Starch is the major metabolite in chestnuts[20], and starch accumulation period of chestnut seeds is the month before maturation. Zhang et al.[21] compared gene expression profiles between two stages (45 vs 75 DAF, days after flowering) of Chinese chestnut and identified two granule-bound starch synthase unigenes exhibiting two-fold higher expression at 75 DAF than at 45 DAF. Two starch branching enzyme isoforms of Chinese chestnut were identified through zymogram analysis. The gene expression of these two CmSBE isoforms increased beginning 64 days after pollination (DAP) and reached the highest levels at 77 DAP[22]. The transcriptomic profiling of Chinese chestnut kernels at 70, 82, and 95 DAF indicated that soluble starch synthase and  $\alpha$ -1,4-glucan branching enzyme genes were actively expressed at 82 and 94 DAF. In addition, starch and sucrose metabolism was significantly enriched in all the comparisons included in the study[19].

Although the regulatory networks of chestnut seed development have been studied, there are few reports on the gene regulatory network governing seed metaxenia during the maturation of chestnut supported by chestnut genome analysis. The first whole-genome sequence of *Castanea mollissima*, completed in 2019[23], provides a promising resource for chestnut functional genomic research. Here, we use ‘Yongfeng 1’, ‘yongren zao’ and ‘Yimen 1’ as male parents to pollinate ‘Yongfeng 1’ as the female parent, and we explore the molecular mechanism of the seed metaxenia effect on starch accumulation. We characterized seed development in chestnut by evaluating nutrient accumulation, by using ELISA to quantify enzyme activity, by examining the expression of key genes through transcriptome profiling and by verifying the expression of genes related to starch synthesis by qRT-PCR. By analyzing the correlation between nutrient changes and seed development, we determined the gene regulatory network controlling the accumulation of starch and soluble sugars in chestnut seed metaxenia and analyzed key genes. We provide abundant genomic resources for chestnut and novel molecular insights into the correlations between the physiological changes and starch accumulation in chestnut metaxenia.

## Results

# Physiological description of the starch accumulation period related to chestnut seed metaxenia

The process of chestnut seed starch accumulation is divided into three stages (S1, S2 and S3) involving changes at the morphological and physiological levels. The seed collection used in this study was obtained from the following different pollination combinations: YFF (self-pollinated ‘Yongfeng 1’), YFR

(‘Yongfeng 1’ × ‘Yongren Zao’) and YFM (‘Yongfeng 1’ × ‘Yimen 1’). As shown in Fig. 1, the seed coloration of YFF, YFR and YFM was similar in the three developmental stages. However, there were significant differences in the seed size of YFF, YFR and YFM at different developmental stages. The seed colors of YFF, YFR and YFM at the S1 (70 DAP), S2 (82 DAP) and S3 (94 DAP) stages were yellow; yellow and brown; and brown, respectively. The seeds of YFF were smaller than those of YFR and YFM at all three stages, which indicated that the metaxenia effect had a significant effect on fruit size (Fig. 1d).

Sugar content, starch content, sucrose synthase (SS) activity and starch synthase (SSS) activity were evaluated in the seeds of chestnut at different developmental stages (Fig. 1a, b, c). There were significant differences in sugar content and SS activity in the seeds of YFF, YFR and YFM at different developmental stages, indicating that metaxenia played a vital role in chestnut seed maturation. The mean soluble sugar content peaked in the S1 stage, then significantly decreased in stage S2 and slightly increased in stage S3 in chestnut seeds. Moreover, the sugar content of YFR was higher than those of YFM and YFF, and YFM exhibited higher SS activity than YFR and YFF, implying that the seeds of YFR ripen more quickly than those of YFM and YFF. Furthermore, the starch content and SSS activity in the seeds of YFF, YFR and YFM exhibited apparent differences, which also indicated that metaxenia played a vital role in chestnut seed maturation. The total starch content, amylose content and SSS activity in the seeds of YFF, YFR and YFM increased with the maturity of seeds. Moreover, the starch content was related to seed color, where the darker the seed color, the higher the starch content was. The sugar content and SS activity were also correlated with the starch content and SSS activity in chestnut seeds. The pollination combinations resulting in a higher sugar content also resulted in a higher starch content and vice versa. Taken together, the results indicated that the sugar content and starch content presented obvious metaxenia effects during the seed starch accumulation period, and the metaxenia effects obviously influenced SS activity and SSS activity .

**Figure 1** Chestnut samples and physiological characteristics at three stages. A: Comparison of sugar contents in three stages. B: Comparison of starch contents in three stages. C: Comparison of enzyme activity in three stages. D: Illustration of seeds at three stages that were included in the study. Note: Lowercase letters indicate significant differences ( $P < 0.05$ ).

## A global transcriptome comparison reveals the relationships between different stages of starch accumulation

To explore the potential molecular mechanism of starch accumulation, global transcriptome analysis was performed in the seeds of chestnut at different developmental stages. RNA was isolated from seeds of YFF, YFR and YFM at S1, S2 and S3 (referred to as YFF-S1, YFF-S2, YFF-S3, YFR-S1, YFR-S2, YFR-S3, YFM-S1, YFM-S2 and YFM-S3), and 27 cDNA libraries (three independent biological replicates for each sample) were constructed. After large-scale sequencing, over 1.14 billion high-quality reads were generated. Approximately 45 million high-quality reads were acquired from each biological replication (Additional file 1: Table S1). The high-quality reads were mapped to the chestnut genome with the TopHat

mapping tool. After treatment with Cufflinks and Cuffmerge, 21,423 genetic loci present in every sample were obtained. Notably, 9999 novel genes were detected. The high-quality reads that were uniquely mapped to the chestnut genome were used for quantification via the fragments per kilobase of transcript length per million mapped reads (FPKM) method. The Spearman correlation coefficients (SCCs) of the biological replicates for each sample ranged from 0.785 to 0.94, except for one replicate of each of YFF-S1 and YFR-S1, which were not used for analysis; these results verified the quality of the obtained reads.

To explore the global differences in transcriptome dynamics during seed starch accumulation in YFF, YFR and YFM, principal component analysis (PCA) of the average FPKM values was performed based on the SCC analysis (Fig. 2, Additional file 1: Figure S2a). The stages showing higher correlations might present more similar transcriptomes and functions. These analyses indicated that there was a higher correlation between similar developmental stages among YFF, YFR and YFM. As expected, the analysis yielded lower correlations during different developmental stages in YFF, YFR and YFM. Although there was a high correlation between stages S1 and S2, stage S3 showed a low correlation with stages S1 and S2. Interestingly, the clustering of YFF and YFR was obviously different. At stage S3, the clustering of YFM was obviously different from that of YFF, which presented a closer correlation with YFR. These results indicated that the seeds of YFR developed faster than those of YFF in the early stages of seed development. The seed development of YFM was similar to that of YFR at the early stage and slower than that of YFR at the later period, which was consistent with the seed morphology of the two

chestnut lines. Taken together, these results indicated that there were major differences in transcription in the three stages of the starch accumulation period in the seeds of each pollination combination.

**Figure 2** Principal component analysis (PCA) of the global transcriptome obtained from three pollination combinations.

## Differentially expressed genes during the chestnut seed starch accumulation period

To better understand gene expression dynamics during the chestnut seed starch accumulation period, differentially expressed genes (DEGs) were identified between these nine samples using the BH algorithm. The differential expression analysis revealed a large number of candidate DEGs according to the following criteria: DEGs were defined according to the ratio of FPKM expression values, a false discovery rate (FDR) < 0.05 and an absolute  $|\log_2(\text{fold-change})|$  value > 0. The genes that were specifically expressed at each stage of the starch accumulation period in the seeds from these three pollination combinations were identified (Fig. 3a). A Venn diagram showed that 1874, 487 and 700 genes were specifically expressed in stage S2 in YFF, YFR and YFM, respectively. A total of 5862, 8873 and 7711 genes were specifically expressed in stage S3 in YFF, YFR and YFM, respectively. Genes that were specifically expressed in different varieties were also identified (Additional file 1: Figure S2a). These results suggested that the highest chestnut seed starch accumulation activity occurred in the S3 stage. Moreover, the most significant starch accumulation was observed in YFR seeds in the S3 stage was, which might lead to early maturity due to metaxenia.

The gene ontology (GO) analysis of DEGs between the seeds of YFF, YFR and YFM at different developmental stages revealed the expression of genes related to the following GO terms: fatty acid metabolic process, organic acid metabolic process, organophosphate metabolic process, monocarboxylic acid metabolic process, amide metabolic process, microtubule motor activity, carbohydrate metabolic process, peptide metabolic process, purine nucleotide metabolic process, ribose phosphate metabolic process, ribonucleotide metabolic process, pyruvate metabolic process and ATP metabolic process (Additional file 1: Figure S2b). These processes are relevant to seed development and starch accumulation. In addition, the GO enrichment analysis of DEGs showed that the S2 stage was characterized by terms related to the cell cycle and growth. The S3 stage was associated with terms related to cell walls, lipid metabolism, secondary metabolism, and protein synthesis. At stage S3, the four most significantly enriched GO terms were amide metabolic process, monocarboxylic acid metabolic process, organic acid metabolic process and protein biosynthesis. In addition, the overlap analysis of GO terms that were significantly enriched in each variety showed that several terms were either specifically enriched in one variety or generally enriched in two varieties (Fig. 3b). For example, GO terms related to carbohydrate derivative metabolic, organophosphate biosynthetic, organophosphate metabolic, and cellular catabolic processes were unique to YFF, while those related to DNA metabolic and amino acid metabolic processes were unique to YFR. GO terms related to microtubule-based, peptide metabolic, translation, monocarboxylic acid metabolic, cellular amide metabolic processes and DNA replication were found in all three varieties. Taken together, these results suggested that certain groups of genes exhibited specific functions in specific stages of the chestnut starch accumulation period. The processes of energy metabolism, organic acid metabolism and translation were presumably activated to promote the synthesis of ATP so that the seeds could obtain more energy during this time. Additionally, ribosomes increase the capacity for protein (peptide) synthesis, and the gluconeogenic pathway enhances sugar accumulation in chestnut seeds. In conclusion, in the early stage of seed development, energy metabolism, organic acid metabolism and translation processes are highly activated, which promotes the synthesis of ATP so that the seeds can obtain more energy. The middle and late stages are characterized by the activation of starch and sugar metabolism, and the accumulation of carbohydrates in chestnut is promoted.

**Figure 3** DEG counts and the top eight GO terms in seeds from three pollination combinations. A: Comparison of DEGs in seeds from different pollination combinations. B: Development-related GO terms in seeds from three pollination combinations.

## Expression patterns of genes responsible for starch accumulation

To analyze the expression profiles of the identified DEGs, 9545, 9966 and 9973 DEGs derived from YFF, YFR and YFM, respectively, were used for cluster analysis with Short Time-series Expression Miner (STEM). Specifically, 5892 DEGs from YFF and 3325 DEGs from YFM were significantly ( $P$  value  $\leq 0.05$ ) aggregated into four profiles, including two downregulated modes and two upregulated modes, and the expression patterns of YFF and YFM were consistent. The 6737 DEGs from YFR were also significantly ( $P$

value  $\leq 0.05$ ) aggregated into four profiles, including two downregulated modes and two upregulated modes, but the YFR expression pattern was inconsistent with those of YFF and YFM (Fig. 4a). GO analysis divided the genes in each profile into three major groups: biological process, molecular function and cellular component. Among the identified cellular components, a large number of upregulated and downregulated DEGs were related to cells, cell parts and organelles. Among biological processes, the two most enriched GO subcategories were cellular processes and metabolic processes. Similarly, for the molecular function category, the two most enriched GO subcategories were catalytic activity and binding.

The expression profiles of these genes were analyzed to explore the expression patterns of key genes encoding six transcription factors (TFs) involved in carbohydrate metabolism in seeds. Most of these genes were highly expressed in YFF in stage S2 (Fig. 4b) and in YFR and YFM in in stage S3. The difference in gene expression found in YFF versus YFR and YFM in the S3 stage was mainly due to the metaxenia effect. The expression of genes encoding glycosyl transferase, carbohydrate phosphorylase, phosphoglucose isomerase, NUDIX and nucleotidyl transferase in the seeds of YFF was increased in stage S2 compared to stage S1 and then decreased in stage S3, which was consistent with the variation trend of SS activity. In thee seeds of YFR, the expression of genes encoding alpha amylase increased with seed maturity, which was consistent with the variation trend of starch contents. However, the expression of genes encoding glycosyl hydrolase and phosphoglucomutase/phosphomannomutase decreased with seed maturity in the seeds of all three pollination combinations. These enzymes are responsible for the biosynthesis of intracellular sugars and participate in the production of energy. These results indicated that the synthesis of soluble sugars, reducing sugars and SS was controlled by regulating the transcriptional activity of the gene encoding glycosyl transferase. However, starch accumulation was related to the phosphorylation of glycosyl hydrolase and glucomutase.

**Figure 4** Expression patterns of genes in the seeds of three pollination combinations at three stages. A: Expression patterns of all genes, B: expression patterns of key genes.

## Hierarchical clustering of all DEGs

Differences in gene expression and regulatory networks are important to elucidate regulatory relationships in seed starch accumulation. To explore the gene regulation network during seed development, coexpressed gene sets of highly expressed genes were identified by weighted gene coexpression network analysis (WGCNA). As shown in the tree diagram (Additional file 1: Figure S3 a), we performed hierarchical clustering of 2345 genes, and a total of 22 different modules, each of which contained 85 to 2222 genes, were identified. Then, each coexpression module was correlated with the seed development stage by Pearson correlation coefficient analysis (Figure S3B). Interestingly, three coexpression modules were uniquely associated with specific stages of seed development, among which the turquoise module, red module and blue module were related to stages S1, S2 and S3, respectively (Additional file 1: Figure S3 b). Moreover, many modules were associated with more than one seed development stage, and some of these modules were related only to the specific developmental stage of a particular variety. For example, the royal blue module and salmon module were highly correlated with stage S3 in YFF and YFM, respectively.

Hub genes related to chestnut seed starch accumulation in different stages were identified by WGCNA (Fig. 5a, b, c). The genes in the purple module were expressed at the highest levels in early stages of seed development (S1). The genes in the pink module exhibited the highest level of expression in stage S2. The genes in the blue module were expressed at the highest levels in stage S3. In the purple module, six hub genes related to protein translation (ribosomal protein and RNA recognition), cell proliferation and sugar biosynthesis were confirmed. Five hub genes involved in translocation, one hub gene encoding transmethylase and one hub gene related to promoting cell proliferation were identified in the pink module. Considering the crucial role of TFs in gene expression and plant development, the expression patterns of TFs were analyzed (Fig. 5d). The RRM and Ras family transcription factors among these hub genes that were closely related to cell proliferation exhibited particularly high transcriptional activity at stage S1. ABC transporter (novel.2300 and novel.2276) and sulfate transporter (CMHBY214244) family genes were highly expressed in stage S2, especially in YFM, suggesting that a large number of genes related to substance transport in seeds were highly expressed under the regulation of TFs. Taken together, the results indicated that large amounts of materials and energy were transported to seeds for anabolic reactions, which was one of the reasons for observing the highest sugar content in seeds in stage S2.

In the blue module, we identified a set of hub genes encoding HSF\_DNA-binding, ACT, Pkinase, and LIM proteins and four transcription factors that were highly expressed in stage S3. The expression of these genes was highest in YFF, followed by YFR and, finally, YFM. The differences in the expression of these genes between the seeds of different pollination combinations were mainly due to the metaxenia effect. These genes play important roles in plant growth, hormone responses and abiotic and biotic stress responses. They might participate in regulating cell senescence and death in chestnut seeds.

**Figure 5** Analysis of key modules based on WGCNA. A: Purple module key genes; B: pink module key genes; C: blue module key genes; D: expression patterns of TFs.

## **Analysis of gene expression related to sucrose and starch metabolism in chestnut seeds**

To identify the major metabolic pathways affecting chestnut seed development, KEGG pathway enrichment analysis of DEGs was performed. The ten most enriched KO terms of each group are listed in Table S2. DEG KO terms related to ribosomal and DNA replication were significantly enriched in the overall process of seed development. The four most enriched DEG KO terms in stage S3 were glycolysis, gluconeogenesis, fatty acid biosynthesis and oxidative phosphorylation. Considering that the expression of the genes related to starch and sucrose metabolism and gluconeogenesis differed significantly in different starch accumulation periods in chestnut seeds, a network of the starch and sugar synthesis pathways in chestnut seeds was constructed (Fig. 6). The network indicated that the starch synthesis pathway and the glycolysis pathway shared the same precursor pool, including g-1-p and g-6-p.

The gene encoding ADP-glucose pyrophosphorylase (EC 2.7.7.27) was highly expressed during the peak period of starch accumulation (S3); the encoded protein catalyzes the conversion of g-1-p to adp-glucose, which is the direct precursor of starch synthesis. The expression of a gene encoding starch granule-

bound starch synthase (EC 2.4.1.242), which is responsible for the synthesis of amylose and amylopectin, was also increased. Moreover, fructose diphosphate aldolase (FBA, EC 4.1.2.13), enolase (EC 4.2.1.11) and pyruvate phosphokinase (EC 2.7.9.1) are significantly regulated by protein phosphorylation. FBA and enolase catalyze reversible reactions in glycolysis. PPDK promotes gluconeogenesis by catalyzing the mutual conversion of pyruvate and PEP. The pyruvate and PEP production steps are initially irreversible steps in glycolysis. The above results indicated that glycolysis is inhibited while gluconeogenesis is enhanced, resulting in the flux of G-1-P and G-6-P to the starch synthesis pathway in chestnut seeds.

**Figure 6** The network of starch and sugar synthesis in chestnut seeds

## Verification By Qrt-pcr

To verify the reliability of the RNA sequencing data, qPCR was used to evaluate the expression profiles of five upregulated and two downregulated genes. The expression trends of the seven selected DEGs were consistent with the transcriptome data (Fig. 7), indicating that the two methods were reliable and were complementary to each other for estimating gene expression.

**Figure 7** Expression of seven selected genes measured by RNA sequencing and qRT-PCR.

## Discussion

Chestnut is a member of the Fagaceae family and is an important economic crop. It has received extensive attention because of its high nutrient content and health-promoting properties [24]. Chestnut fruit development is accompanied by starch accumulation during the ripening process, in which sugar synthesis and decomposition pathways play a key role [21]. Studies have shown that different varieties of chestnut are characterized by different processes during development and maturation [19]. It has been demonstrated that the levels of N, sugar, starch, fat and K are significantly higher under cross-pollination than self-pollination treatments during ovary development in chestnut [25]. Our results are consistent with previous studies. In this study, the metaxenia effect was shown to lead to differences in the accumulation of starch in seeds produced from different pollination combinations in chestnut, and cross-pollination resulted in a higher starch content than self-pollination. In addition, our research showed that metaxenia affects the size and color of nuts, which is consistent with findings in other species, such as hazelnut [26] and tomato [27].

Glycosyltransferases (GTFs) are enzymes (EC 2.4) that establish natural glycosidic linkages. They catalyze the transfer of saccharide moieties from an activated nucleotide sugar (also known as the "glycosyl donor") to a nucleophilic glycosyl acceptor molecule, the nucleophile of which can be oxygen carbon, nitrogen, or sulfur based [28–29]. In our study, the expression levels of genes encoding the glycosyl hydrolase family and phosphoglucomutase/phosphomannomutase were shown to play an important role in the seeds from all three pollination combinations. Glycosyl hydrolase controls the

hydrolysis of glycosyl groups and participates in energy metabolism, and the change trend of this enzyme was opposite that of the accumulation of starch. The results showed that the most abundantly expressed unigenes according to total RPKM values at 45 DAF and 75 DAF included GBSS (EC 2.4.1.186). The RPKM of GLG (EC 2.4.1.242) in the transcriptome was extremely low at 45 DAF and undetectable at 75 DAF. The GBSS and GLG gene families were identified from glycosyltransferase categories according to the KEGG protein database, which indicated that GTFs also play a very important role in the development of chestnut seeds[21]. However, another study indicated that the SuS (EC2.4.1.13) and SEB (EC 2.4.1.18) genes were actively expressed at 82 and 94 DAF and that SP (EC 2.4.1.1)-encoding unigenes were significantly downregulated at 94 DAF. GTFs seem to be less important in the process of seed development[19]. A gene encoding ADPase (EC 2.7.7.27) was shown to be highly expressed in the peak period of starch accumulation (94 DAF), consistent with the findings of Li et al.[19].

In addition to the enzymes closely related to sugar metabolism, the results of GO and KEGG enrichment analysis showed specificity in the seeds of different pollination combinations and in different developmental stages. In stage S3 (94 DAF), amide metabolism, monocarboxylic acid metabolism, organic acid metabolism, and protein synthesis were significantly enriched. The seeds of the different pollination combinations were also characterized by many specific enriched terms. In YFF, GO terms such as carbohydrate derivative metabolic process, organophosphate biosynthetic process, organophosphate metabolic process, and cellular catabolic process were significantly enriched; in YFR, DNA metabolic process and amino acid metabolic process were specifically enriched. To fully understand the biosynthesis of starch and its regulation, it will be very important to study these abundant pathways and related genes in the future.

To more systematically analyze the differences in gene expression and regulatory networks during the development of chestnut kernels, we identified coexpressed gene sets by WGCNA. At 70 DAF, we identified six genes at the core position, most of which were related to translation, cell proliferation, and sugar synthesis. At 82 DAF, a module with the highest correlation was identified, and this module exhibited transport-related activities. At this stage, both SS and sugar metabolism in chestnut kernels showed high activity, which might be one of the reasons for the high transport activity observed. In the last stage, we identified transcription factors related to plant growth, hormone regulation and biotic and abiotic stress responses, such as ACT and LIM TFs, which may be involved in the regulation of cell senescence and death in chestnut seeds.

## Methods

### Plant materials and sampling

The experimental site was located in Weidi Township, Yongren County ((Lat.25°51'~26°30' N, 101°14'~101°49' E, Alt. 1530 ~ 1700 m)), Yunnan Province. 'Yongfeng 1' (Yunnan Province certification local fine variety) was selected as the female parent. The following pollination combinations were established: 'Yongfeng 1' × 'Yongfeng 1' (YFF), 'Yongfeng 1' × 'Yongren Zao' (YFR), and 'Yongfeng 1' ×

'Yimen 1' (YFM). Four sampling directions (east, west, north and south) were selected for performing the pollination combinations in the above groups. At the early stage of the germination of female flowers (2018/04/29), approximately 80 samples from each pollination combination were bagged. At the flowering stage of male flowers, fresh pollen was collected for direct pollination (2018/05/19) and was removed from the bag prior to thorn ball swelling (2018/05/26).

Description of chestnut varieties: 'Yongfeng 1' is a real-breeding variety in Yongren County, Yunnan Province, with an average single seed weight of approximately 16 g. 'Yongren Zao' is also a real-breeding variety in Yongren County, Yunnan Province, with an average single-grain weight of 12 g, in which the mature period occurs in late July. 'Yimen 1' is a well-approved variety in Yunnan Province. its fruit ripening period is in mid-August, and it presents an average single-grain weight of 14.51 g.

Starting approximately 20 days before maturation, samples were collected every 11 days, and the last samples obtained consisted of mature fruits with thorn ball dehiscence. Chestnut seeds were collected from 18 pollination trees at three developmental stages (70, 82 and 94 DAP) (Fig. 1). The endosperm was removed from the seed and divided into three equal parts. Then, two of the parts were wrapped with silver paper and immediately stored in liquid nitrogen until being used for RNA extraction, and the third part was brought directly to the laboratory, where it was dried and ground.

## **Physiological and biochemical indexes of sugar and starch**

The seed endosperms were dried in a 60 °C oven. The soluble sugar content was determined via the colorimetric anthrone method [30]. The reducing sugar content was determined by direct titration [31], and amylopectin and amylose contents were determined using a microanalytical method [32]. SSS and SS levels were tested by double-antibody sandwich ELISA [33].

## **Gene expression analyses**

Total RNA from each seed sample was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, USA), and an Agilent 2100 system was used to detect RNA. Sequencing libraries were constructed by using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) according to the manufacturer's recommendations. mRNA was enriched with mRNA capture beads and then fragmented. First-strand cDNA was synthesized using random hexamers, followed by synthesis of the double-stranded cDNA. After end repair, an A tail and an adapter were added, and cDNA fragments 150–200 bp in length were selected. Sequencing libraries were generated via PCR. Then, we measured the quality of the libraries. The libraries were sequenced using a HiSeq 2000 sequencer (Illumina, San Diego, CA, USA), and 150 bp paired-end reads were generated. Adapters and low-quality reads were removed from the raw data to obtain the clean data. We also measured the Q20 and Q30 values of the clean data. The high-quality cleaned reads were mapped to the Chinese chestnut reference genome[22] using HISAT 2 with default parameters. The unique mapping read counts were normalized to FPKM values by Cufflinks (v2.0.2) to obtain the relative expression levels for each sample.

PCA was performed to show the correlations between biological replicates using an R package. We performed differential expression analysis based on Cufflinks. Genes with adjusted P-values (q value)  $\leq$  0.05 and an absolute log<sub>2</sub> (fold change) value were identified as DEGs. To observe the gene expression patterns in seeds in different stages, the DEGs were grouped by hierarchical clustering according to their level of expression using the heatmap R package.

## Functional enrichment analysis

We performed GO enrichment analysis of the DEGs according to GO:TermFinder (Boyle et al., 2004). Significantly enriched GO terms for DEGs compared with the genomic background were identified using a hypergeometric test. A corrected p value  $<$  0.05 was set as the threshold for significantly enriched GO terms. Similarly, we identified the statistical enrichment of DEGs in KEGG pathways using KOBAS software (v2.0). MapMan (v.3.5.1R2) was used to visualize the expression patterns of DEGs.

## Weighted gene coexpression network analysis

To detect patterns of gene connectivity, we analyzed the data via WGCNA with the WGCNA package (v1.51) in R. Genes with a low coefficient of variation of the averaged RPM ( $CV < 1$ ) among all samples were discarded, and the remaining 12,662 genes were used for the analysis.

First, the scale-free gene network condition should be satisfied before conducting WGCNA. Moreover, it is necessary to define the correlation matrix of gene coexpression and adjacency function. A soft threshold (power) of five was chosen in this study. Then, the dissimilarity measurements of different nodes were calculated, and a hierarchical clustering tree was built based on these data. Different dendrogram branches represent various modules. We draw the heatmap via the module eigengene that is used to show the expression pattern in every sample. Therefore, we can study a module that is significantly related to further research via heatmap data. The functional significance of modules was evaluated through GO and KEGG enrichment analysis according to a previous method. The coexpression networks for selected modules of genes were visualized by Cytoscape (version 3.1.0).

## qRT-PCR identification

The expression patterns of 14 unigenes encoding SSS and SS in developing Chinese chestnut seeds were studied by qRT-PCR. Total RNA was extracted using a Qiagen RNeasy Mini Kit (Qiagen Inc., Valencia, CA) and was reverse transcribed into cDNA by using random primers. The internal control was 18S rRNA[34]. The  $2^{-\Delta\Delta CT}$  method was employed to analyze the data [35].

## Abbreviations

DEGs: differentially expressed genes; FPKM: fragments per kilobase of transcripts per million mapped fragments;

## Declarations

## Acknowledgments

We are very grateful to the editor and reviewers for critically evaluating the manuscript and providing constructive comments for its improvement.

## Funding

This work was supported by a grant from the Science and Technology Major Project of Guangxi(Guike AA17204058-03) and the Independent project of Key Laboratory of Cultivation and Utilization of Guangxi Characteristic Economic Forest(16-A-0401).

## Availability of data and materials

Sequencing data are available via the NCBI Bioproject PRJNA540079.

## Author Contributions

LS designed and performed the experiments, analyzed the data, drafted the paper, and prepared figures and tables. SZ and ZH conceived the study and acquired funding for the project. ZQ and TL's contributions included advising the project and revising the manuscript. LW assisted with the experiments.

## Ethics approval and consent to participate

This study did not directly involve humans or animals. Chestnut seeds were collected from the chestnut planting base of Southwest Forestry University, Weidi Township, Yongren County (Lat.25°51'~26°30' N, 101°14'~101°49' E, Alt. 1530~1700 m), China. The research conducted complied with all institutional and national guidelines.

## Consent for publication

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

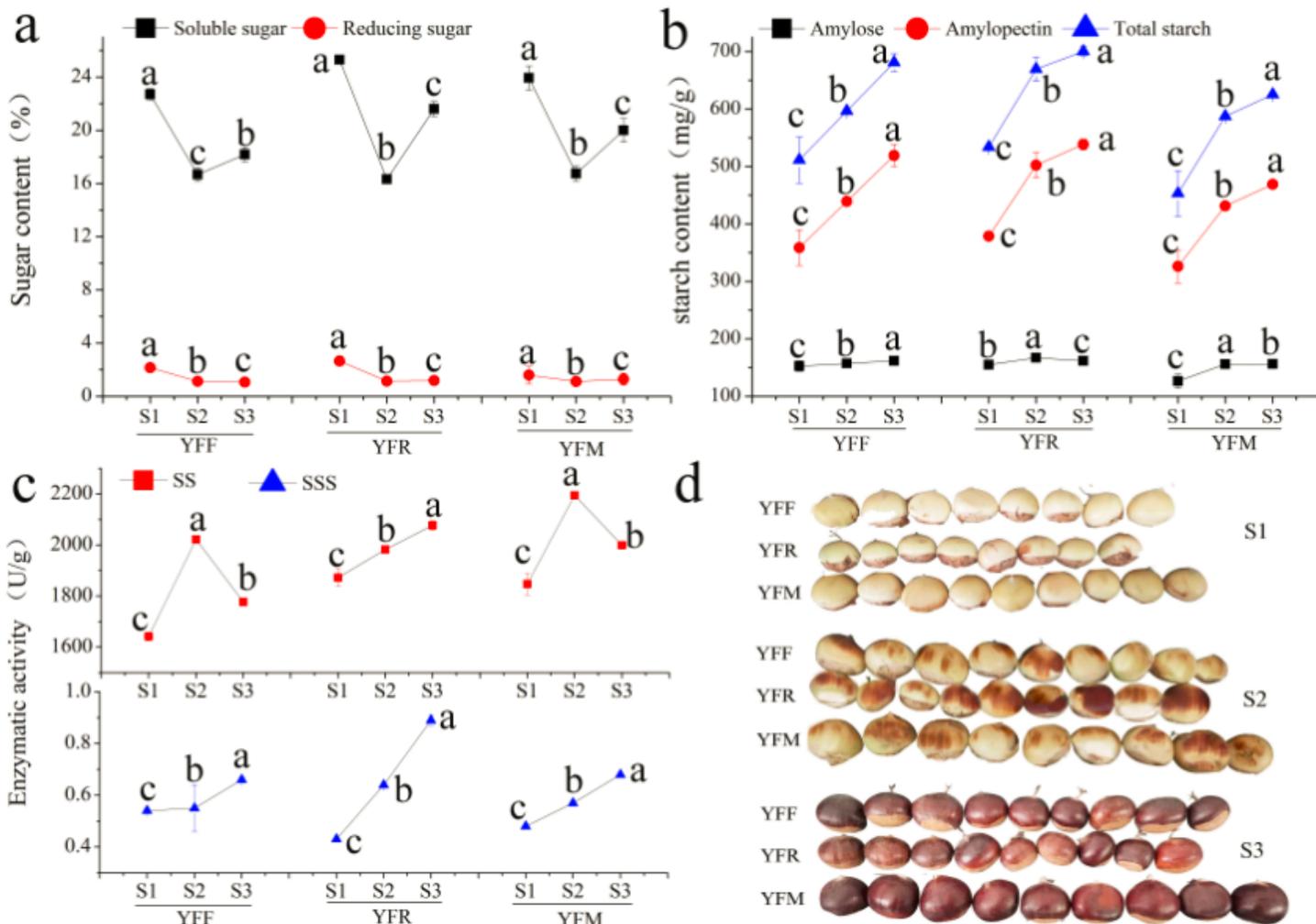
## References

1. Shi Z, Xia L I. Stigmatic morphology of Chinese chestnut (*Castanea mollissima* Blume).[J]. Hortscience, 2010, 45(6):981–983.
2. De Vasconcelos, Maria Do Carmo Barbosa Mendes, Bennett R N, Rosa E A S, et al. Primary and Secondary Metabolite Composition of Kernels from Three Cultivars of Portuguese Chestnut (*Castanea sativa*, Mill.) at Different Stages of Industrial Transformation[J]. Journal of Agricultural and Food Chemistry, 2007, 55(9):3508–3516.

3. Shi Z, R. Stösser. Reproductive Biology of Chinese Chestnut (*Castanea mollissima* Blume)[J]. *European Journal of Horticultural Science*, 2005, 70(2):96–103.
4. Borges O, Gon Alves B, José L. Soeiro de Carvalho, et al. Nutritional quality of chestnut (*Castanea sativa* Mill.) cultivars from Portugal[J]. *Food Chemistry*, 2008, 106(3):976–984.
5. Cruz B R, Abra O A S, Lemos, André M, et al. Chemical composition and functional properties of native chestnut starch (*Castanea sativa* Mill)[J]. *Carbohydrate Polymers*, 2013, 94(1):594–602.
6. Wang, Q., Su, S., Zhao, D., Kou, Y., Study on Effect of Different Varieties Configuration upon Yield of Yan Mountain Chestnut. *American Journal of Plant Sciences*, 2012, vol.3, pp. 738–744.
7. Wang Jing, Guo Sujuan. Effects of pollination with different pollen on the setting characteristics and fruit quality of Chinese chestnut [J]. *Forestry Science of Taiwan*, 2018, 33 (4): 263–275
8. Focke Wo. Die pflanzen mischung e: ein beitr ag zur biolog ie der g ewa ^chse[ M ]. Berlin: Bor nt raeger, 1881:510–518.
9. Denny J O. Xenia includes metaxenia[ J]. *Hor t Science*, 1992, 27( 7) : 722–728.
10. Olfati J A, Sheykhtaher Z, Qamgosar R, et al. Xenia and Metaxenia on Cucumber Fruit and Seed Characteristics[J]. *International Journal of Vegetable Science*, 2010, 16(3):243–252.
11. Daulta, B. S, Chauhan, K.S. Metaxenia Studies on Some Berry and Seed Characters in Grapes (*Vitis Vinifera* L.)[J]. *Indian Journal of Horticulture*, 1984.
12. Nebel B R. METAXENIA IN APPLES[J]. *Journal of Heredity*, 1936(9).
13. Bulant, C., et al. "Xenia effects in maize with normal endosperm." (2000): 182–189.
14. Ehlenfeldt M K. Investigations of metaxenia in northern highbush blueberry (*Vaccinium corymbosum* L.) cultivars[J]. *Journal of the American Pomological Society*, 2003, 57(1): 26.
15. Zhao zhiheng, Yang liu, shi zhuogong, et al. Effects of pollen anaesthetic effect of chestnut on inner quality of nuts [J]. *Research on economic forest*, 2012, 30(4):9–12.
16. Yang liu, zhao zhiheng, shi zhuogong. Effects of anaesthesia effect of chestnut pollen on sugar content in fruit [J]. *Journal of northwest forestry university*, 2012, 27(6):75–77.
17. Zhou jin. Effects of pollen anaesthesia and soil physical and chemical properties on chestnut fruit quality [D]. *Beijing forestry university*, 2010.
18. Feng Z, Su-Juan G, Jing W, et al. Determination the Change of Main Trace Elements in the Ovary with Self- and Cross-Pollination of Chinese Chestnut by ICP-MS[J]. *Guang Pu*, 2015, 35(10):2867–2871.
19. Li S, Liang H, Tao L, et al. Transcriptome Sequencing and Differential Expression Analysis Reveal Molecular Mechanisms for Starch Accumulation in Chestnut[J]. *Forests*, 2020, 11(4):388.
20. Liang, L.; Xu, J.; Wang, G.; Ma, H. Relationship between starch pasting, amylose content and starch granule size in different Chinese chestnut variety groups. *Sci. Agric. Sinica* 2009, 42, 251 – 260.
21. Zhang, L.; Lin, Q.; Feng, Y.; Fan, X.; Zou, F.; Zeng, X.CH.; Cao, H.P.; Yuan, D.Y. Transcriptomic identification and expression of starch and sucrose metabolism genes in the seeds of Chinese chestnut (*Castanea mollissima*). *J Agric Food Chem*. 2015b, 63, 929–9.

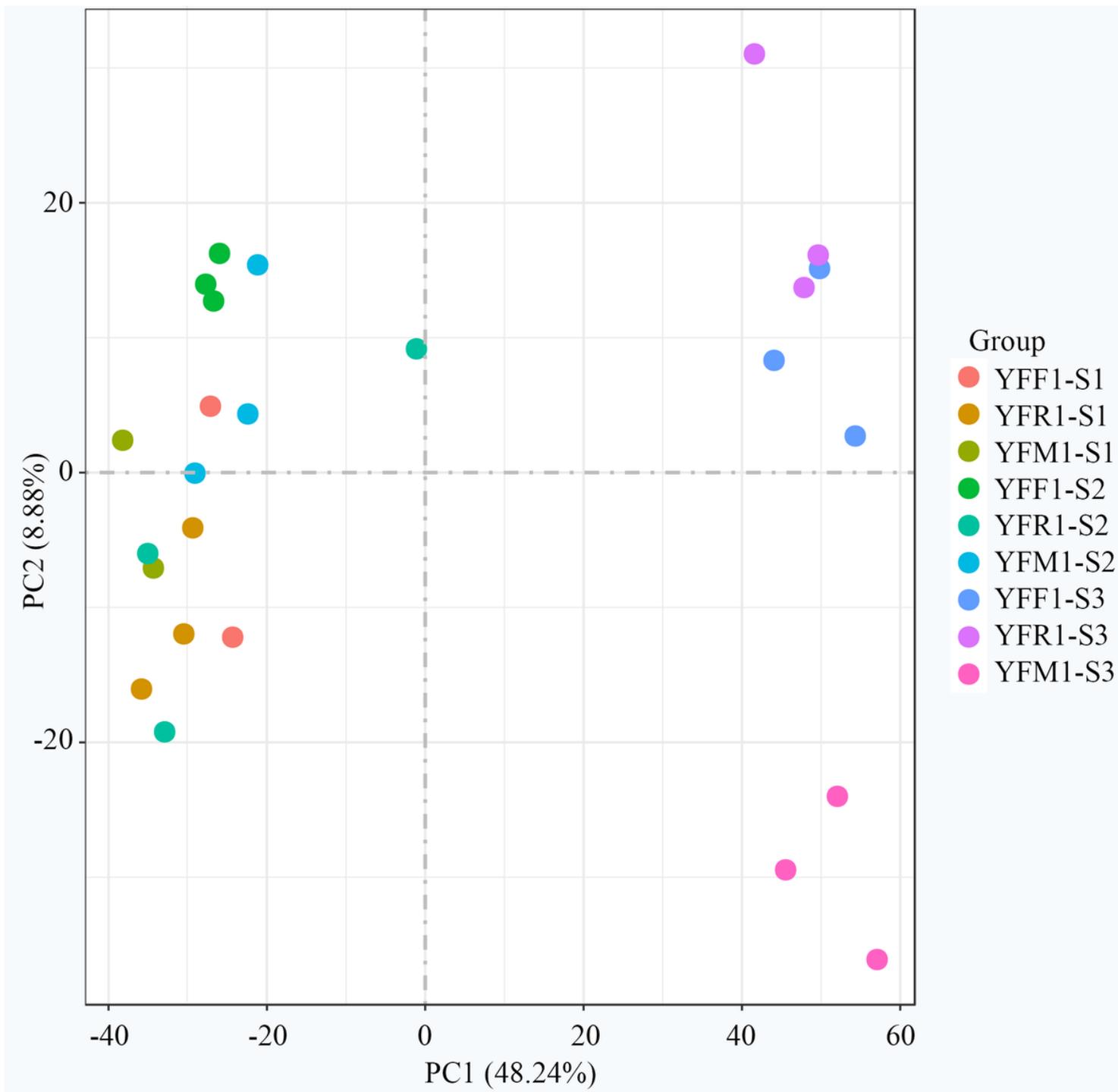
22. Chen, L.; Lu, D.; Wang, T.; Li, Z.; Zhao, Y.; Jiang, Y.; Zhang, Q.; Cao, Q.; Fang, K.; Xing, Y.; Qin, L. Identification and expression analysis of starch branching enzymes involved in starch synthesis during the development of chestnut (*Castanea mollissima* Blume) cotyledons. *PLOS ONE* 2017, 12, e017779242.
23. Yu X, Yang L, Qing Z, et al. Hybrid de novo genome assembly of Chinese chestnut (*Castanea mollissima*)[J]. *Gigaence*(9):9.
24. Zhang Y. Comparison of the transcriptomes of American chestnut, and Chinese chestnut, in response to the chestnut blight infection[J]. *Bmc Plant Biology*, 2009, 9.
25. Xiao-na, Zhang, Ya-qin, et al. Effects of Different Pollination Treatments on Nutrition Changes of the Ovary in Chinese Chestnut (*Castanea mollissima* Blume)[J]. *Advance Journal of Food Science & Technology*, 2015.
26. Hüseyin rfan Balk, Beyhan N. Xenia and metaxenia in hazelnuts: Effects of pollinizer cultivars on nut set and nut characteristics of some hazelnut cultivars[J]. 2020.
27. Piotta FA, Batagin-Piotta KD, de Almeida M, et al (2013). Interspecific xenia and metaxenia in seeds and fruits of tomato[J]. *Scientia Agricola*, 70 (2):102–107.
28. Zhang S, Dong J G, Wang T, et al. High level accumulation of  $\alpha$ -glucan in maize kernels by expressing the *gtfD* gene from *Streptococcus mutans*[J]. *Transgenic Research*, 2007, 16(4):467–478.
29. Frederic P. Miller, Agnes F. Vandome, John McBrewster. *Glycosyltransferase*[J]. alphascript publishing, 2010, volume 17(4):e33-e33(1).
30. Oleksyn, J.; Zytowski, R.; Karolewski, P.; Reich, P.B.; Tjoelker, M.G. Genetic and environmental control of seasonal carbohydrate dynamics in trees of diverse *Pinus sylvestris* populations. *Tree Physiol.* 2000,20, 837–847.
31. Alexander, P.W.; Hartati, R.D.; Curtin, J. Automated potentiometric end point determination in the lane Eynon titration of reducing sugars. *Electroanalysis*. 1989, 1, 263–269.
32. Jones, A.; Urban, J.; Copíková, J. A micro-analytical method for the determination of starch and amylose/amylopectin content in pea seeds. *Biol Plant*. 1999, 42, 303–308.
33. Blundell M J, Phongkham T, Gale K R. A novel enzyme-linked immunosorbent assay for the detection of Wx-B1 null wheat lines using a monoclonal antibody and its application for the detection of marker heterogeneity within commercial cultivars[J]. *Journal of Cereal Science*, 2010, 52(2):187–191.
34. Kuchipudi S V, Tellabati M, Nelli R K, et al. 18S rRNA is a reliable normalisation gene for real time PCR based on influenza virus infected cells[J]. 2012, 9(1).
35. Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta C(T)}$  method. *Methods*, 2001, 25(4):402 – 408.

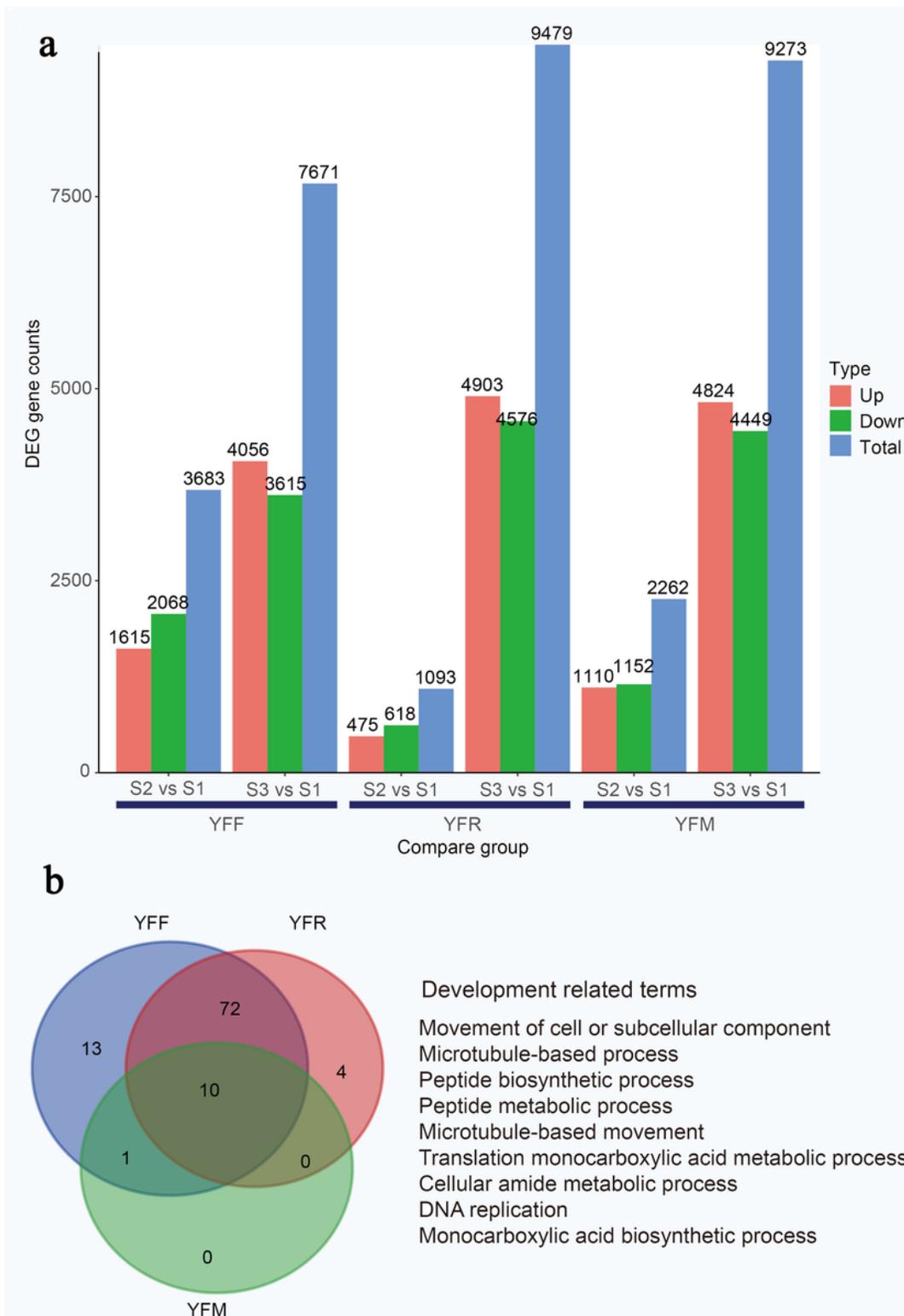
## Figures



**Figure 1**

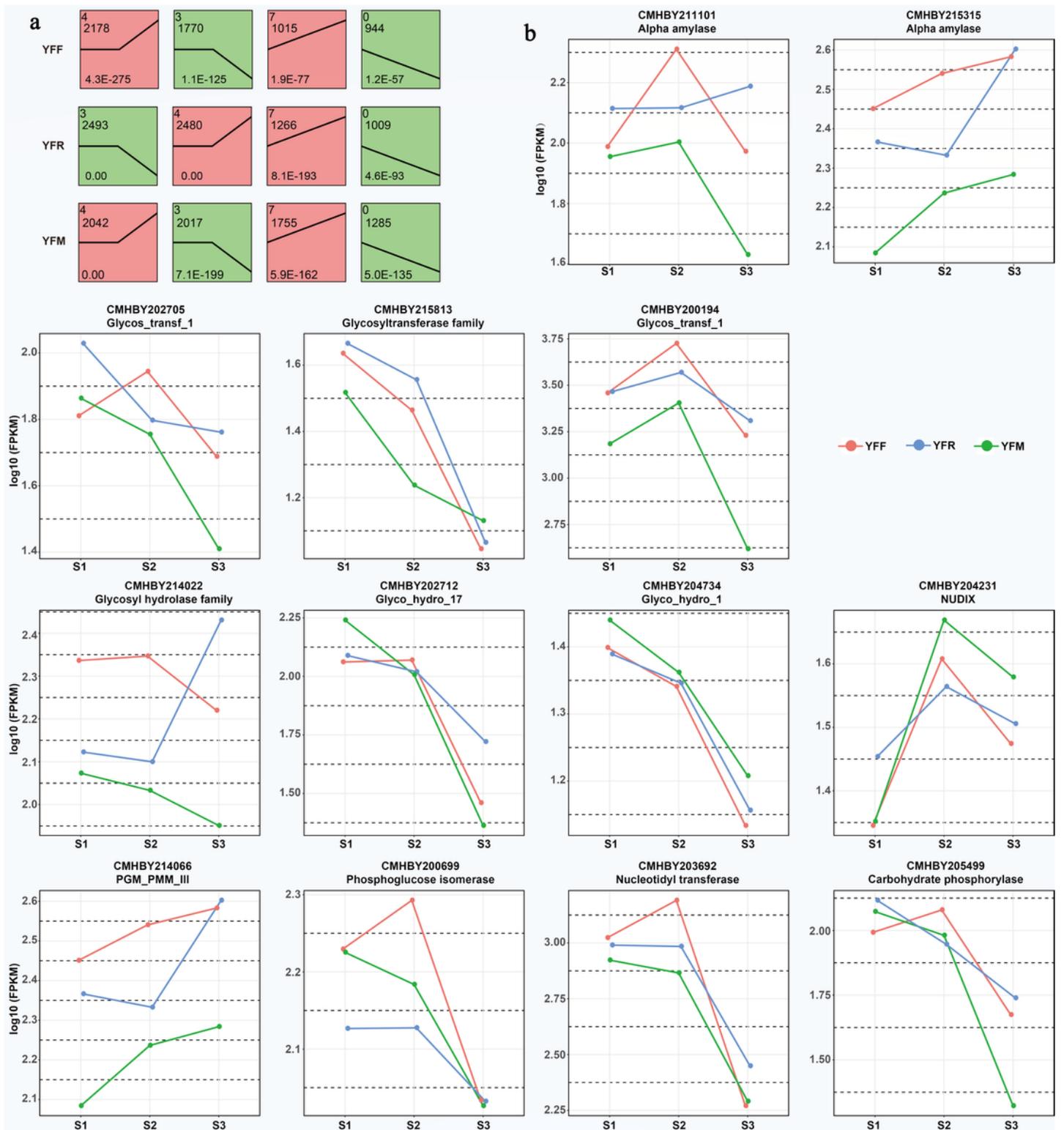
Chestnut samples and physiological characteristics at three stages. A: Comparison of sugar contents in three stages. B: Comparison of starch contents in three stages. C: Comparison of enzyme activity in three stages. D: Illustration of seeds at three stages that were included in the study. Note: Lowercase letters indicate significant differences ( $P < 0.05$ ).





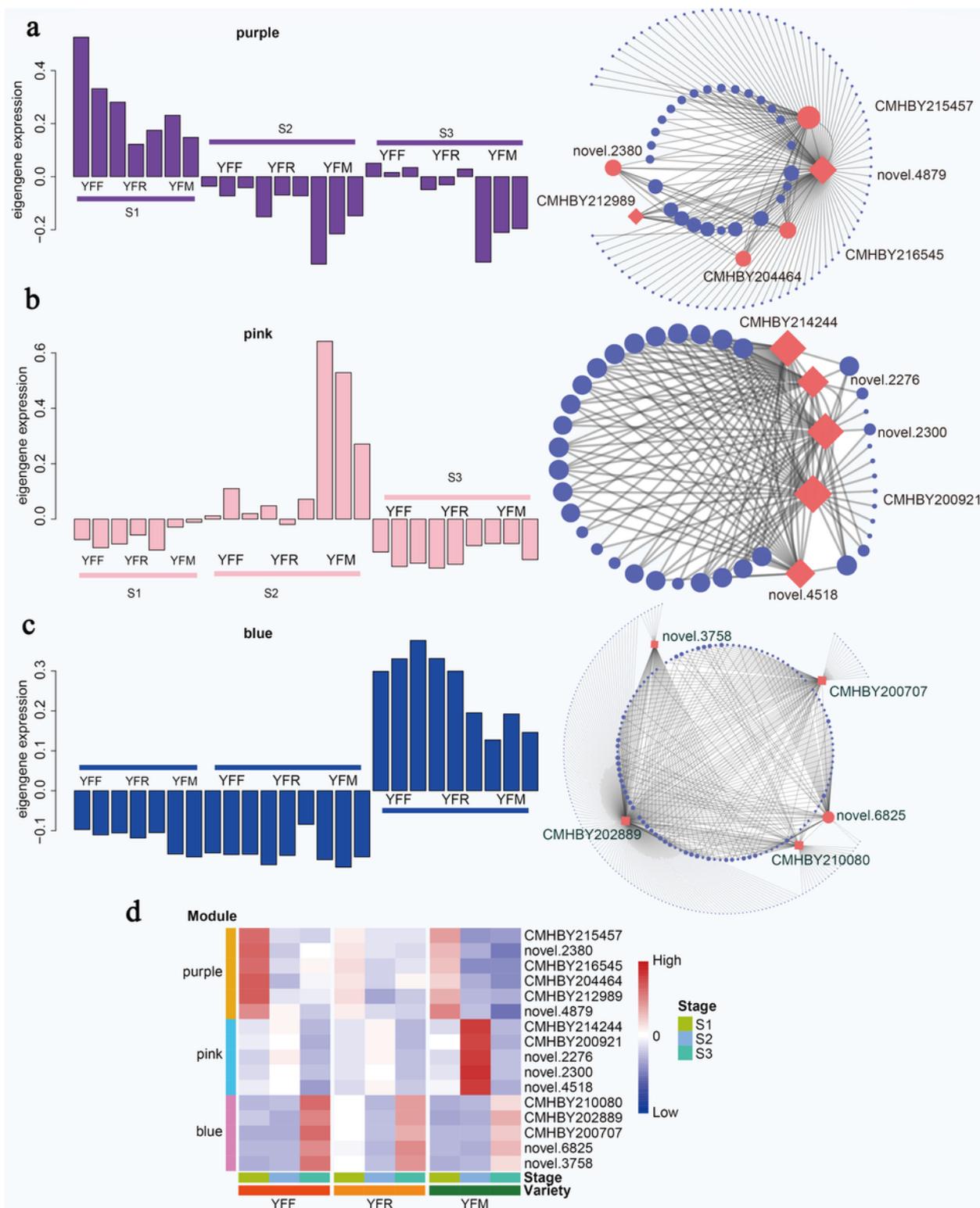
**Figure 3**

DEG counts and the top eight GO terms in seeds from three pollination combinations. A: Comparison of DEGs in seeds from different pollination combinations. B: Development-related GO terms in seeds from three pollination combinations.



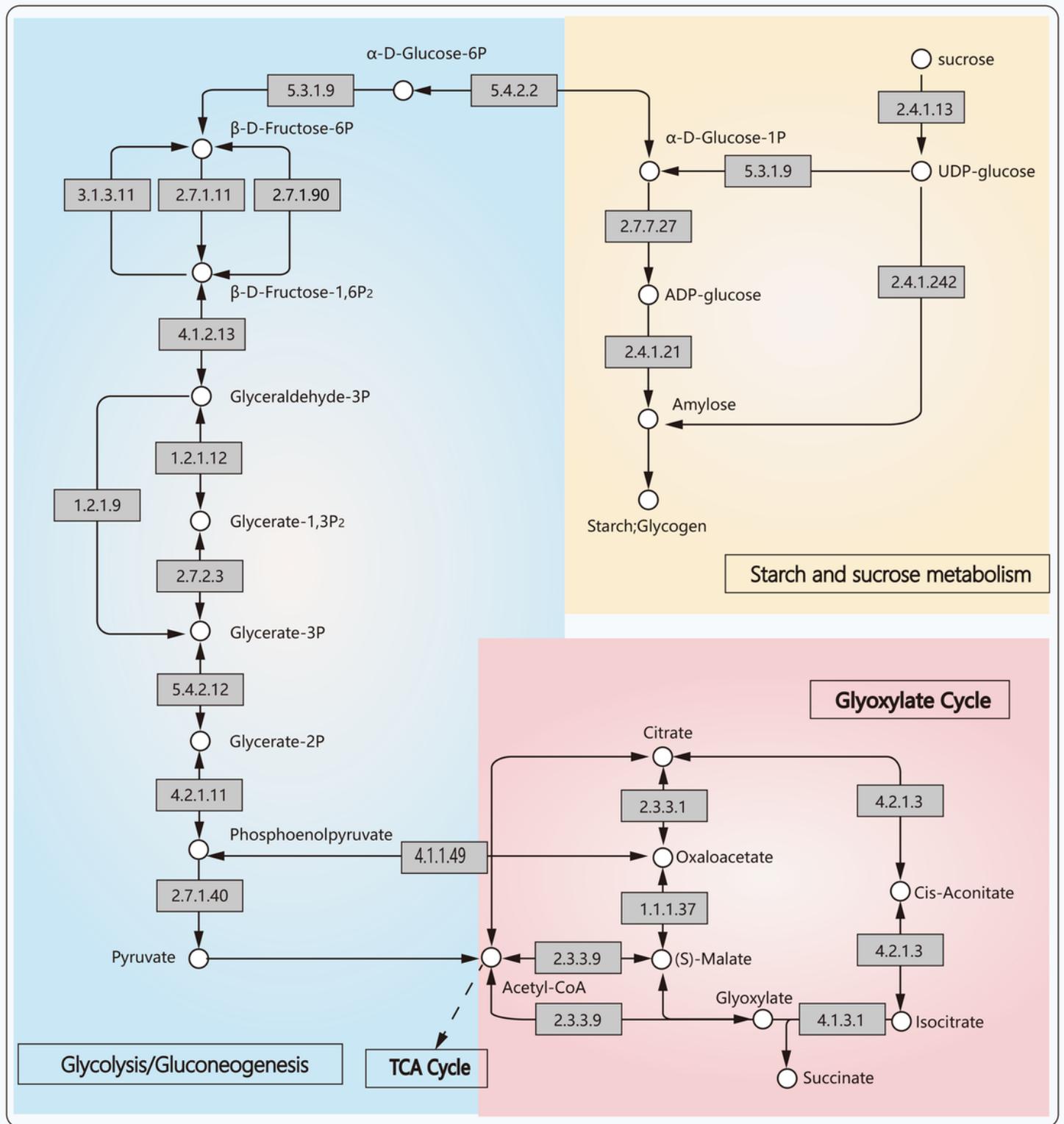
**Figure 4**

Expression patterns of genes in the seeds of three pollination combinations at three stages. A: Expression patterns of all genes, B: expression patterns of key genes.



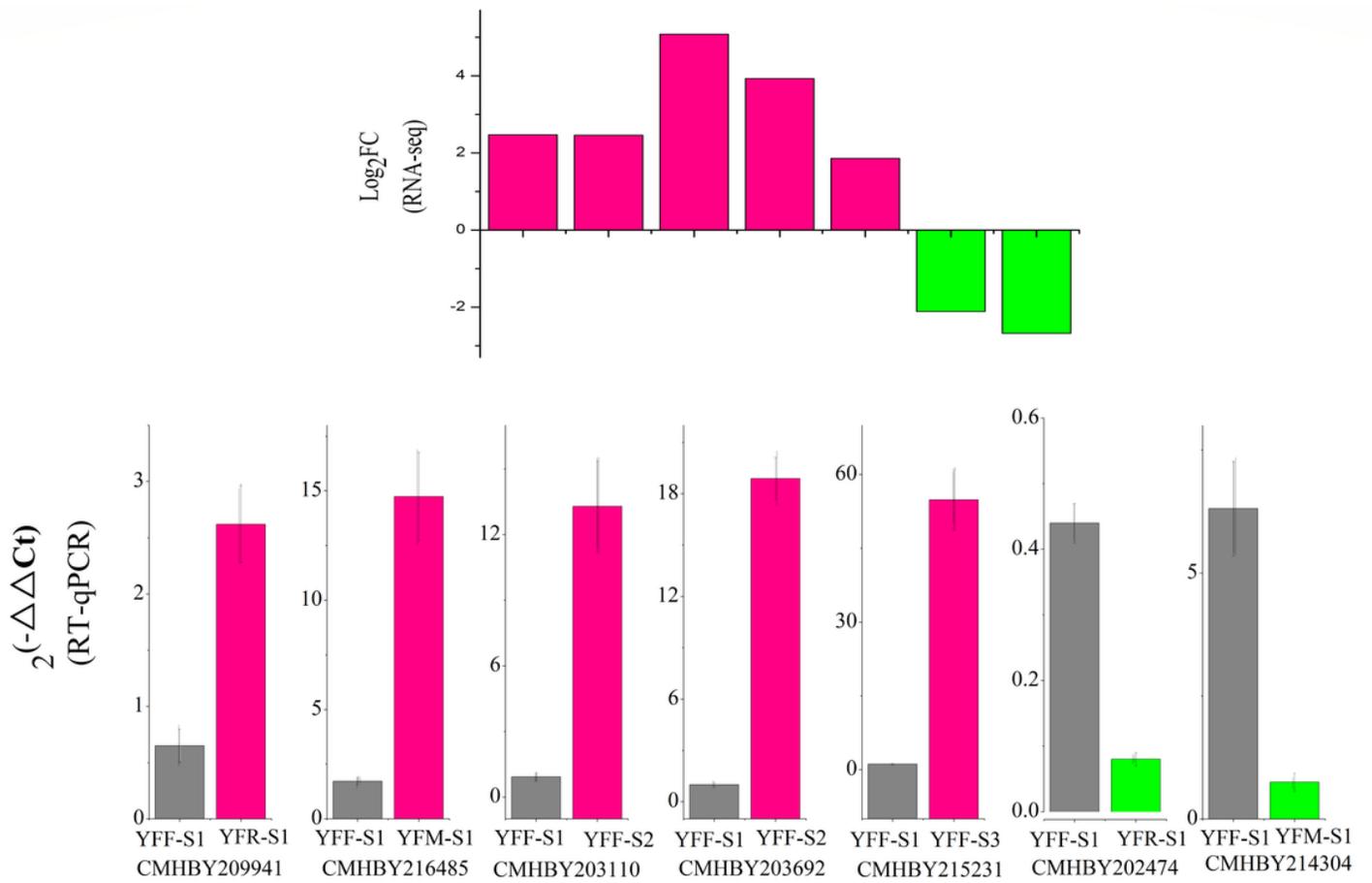
**Figure 5**

Analysis of key modules based on WGCNA. A: Purple module key genes; B: pink module key genes; C: blue module key genes; D: expression patterns of TFs.



**Figure 6**

The network of starch and sugar synthesis in chestnut seeds



**Figure 7**

Expression of seven selected genes measured by RNA sequencing and qRT-PCR.

## Supplementary Files

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

- [figS1.tif](#)
- [figS2.tif](#)
- [figS3.tif](#)
- [TableS1.xlsx](#)
- [TableS2.xlsx](#)