Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Transcriptome and Co-expression Network Analysis Reveals Key Modules and Hub Genes Controlling Husk Number in Maize

Xitong Xu
Shandong Agricultural University
Dusheng Lu
Shandong Agricultural University
Yancui Wang
Shandong Agricultural University
Yanbin Han
Shandong Agricultural University
Haixiao Su
Shandong Agricultural University
Lifan Shi
Shandong Agricultural University
Cuixia Chen
cxchen@sdau.edu.cn
Shandong Agricultural University

Research Article

Keywords: Maize (Zea mays L.), Husk number, WGCNA, Hierarchical clustering, Transcriptome, Candidate genes

Posted Date: July 29th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10440596/v1

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

Additional Declarations: No competing interests reported.

Transcriptome and Co-expression Network Analysis Reveals Key Modules and Hub Genes Controlling Husk Number in Maize

Xitong Xu[1], Dusheng Lu[1], Yancui Wang[1], Yanbin Han[1], Haixiao Su[1], Lifan Shi[1], Cuixia Chen*

[1]College of Agronomy, Shandong Agricultural University, Tai'an 271018, China.

*Corresponding author. Email: cxchen@sdau.edu.cn

Abstract

Background: Husk number is a key agronomic trait in maize that affects grain dehydration rate and yield, yet its regulatory mechanisms remain largely unknown. This study aimed to systematically identify the gene modules and hub genes controlling husk number using an integrated transcriptome approach.

Results: RNA-seq was performed on ear tissues from two maize inbred lines with extreme husk numbers (LJ311, ~20 husks; PHG35, ~5 husks) at four developmental stages (V7–V10). A total of 12,605 husk-specific genes were identified. Hierarchical clustering combined with differential expression analysis revealed a late-stage-specific gene cluster (Cluster 3, 44 genes) that was strongly upregulated at V8–V10 in the high-husk line. WGCNA identified three co-expression modules significantly correlated with husk number: MEblue and MEbrown (positively) and MEturquoise (negatively). Cluster 3 shared only seven genes with MEblue and was enriched in the TCA cycle, DNA replication, and glycerophospholipid metabolism. By intersecting hub genes (|kME| > 0.9) with high-confidence differentially expressed genes (|log2FC| > 2), two final candidate gene sets

were obtained. The upregulated candidates were enriched in cell cycle, brassinosteroid signaling, and phospholipid metabolism, whereas the downregulated candidates were enriched in MAPK signaling, glycolysis, and cell wall polysaccharide degradation.

Conclusion: The results suggest that increased husk number is driven by a BR–MAPK–cell cycle regulatory program, with an opposing mechanism in the low-husk line. This study provides an integrated regulatory network and valuable candidate genes for future functional validation and molecular breeding.

Keywords: Maize (*Zea mays* L.), Husk number, WGCNA, Hierarchical clustering, Transcriptome, Candidate genes

Background

Maize (*Zea mays* L.) is one of the most important food crops in the world. Its global production reached 1.287 billion tons in 2025 (https://www.fao.org/). Over recent decades, mechanical harvesting of maize has greatly reduced labor intensity and saved labor costs. As a result, achieving full mechanization of maize production has become an urgent task[1, 2]. Among the many technical challenges, mechanized grain harvesting remains the final bottleneck. Currently, the grain moisture content of most maize varieties in China reaches 30%–40% at harvest. This value is far higher than the standard suitable for mechanized harvesting, which is 15%–25% [3, 4]. As a result, the mechanized grain harvesting area for maize in China is less than 5%. High grain moisture not only seriously limits harvesting efficiency but also increases drying and storage costs. In addition, it leads to a higher rate of kernel breakage during mechanical harvesting. Under warm and humid conditions, it can also cause problems such as kernel mold, fungal infection, and ear sprouting [5-7]. Therefore,

increasing the kernel dehydration rate (KDR) during late developmental stages and reducing grain moisture content at harvest have become two of the most important objectives in modern maize breeding programs [8].

Grain dehydration rate is affected by many factors. Among them, husk morphology plays a key role. A husk is a modified leaf that encloses the maize ear. From the perspective of plant developmental biology, it is a protective structure produced by axillary meristems before floret formation [9, 10]. The developmental pattern of husk leaves is very different from that of normal leaves. In terms of morphogenesis, the husk sheath originates from lateral meristems and is homologous to the leaf sheath that surrounds the stem. The husk blade, Meanwhile, develops from the further extension of the top of the husk sheath [11]. Studies have shown that husk area is significantly and positively correlated with grain moisture content at harvest. Maize varieties with a smaller husk area tend to have a faster grain dehydration rate [5]. Too many husk leaves create a sealed micro-environment around the ear and restrict air movement. As a result, grain water loss is greatly slowed down. In contrast, several husk traits can help speed up grain dehydration. These traits include fewer husk layers, thinner husk leaves, loose husk coverage, and early husk senescence [12, 13]. Many studies have confirmed that husk number, husk length, and husk coverage are all significantly and positively correlated with grain moisture content at harvest [14, 15]. Therefore, improving husk traits through genetic approaches has become a key goal in maize breeding. The aim is to maintain sufficient protection while also promoting efficient grain dehydration, so that the new varieties will be more suitable for mechanized harvesting.

In recent years, quantitative trait locus (QTL) mapping and genome-wide association studies (GWAS) in different maize populations have identified several QTLs and candidate genes related

to husk number [16-19]. However, because husk traits have complex quantitative genetic characteristics and single-gene approaches have limitations, the global gene regulatory network and core functional modules controlling husk number determination are still not fully understood. Transcriptome sequencing (RNA-seq) combined with computational biology methods has become an important strategy for dissecting the molecular basis of complex agronomic traits [20]. Differential expression analysis can identify genes that are significantly differentially expressed between different genotypes or developmental stages. This step provides an initial pool of candidate genes for downstream analyses [21]. Unsupervised hierarchical clustering, in turn, groups genes based on the similarity of their expression patterns. It can reveal naturally occurring gene sets with potential functional coherence [22]. Weighted gene co-expression network analysis (WGCNA) further organizes genes into co-expression modules. It then correlates these modules with external phenotypic traits. This allows the identification of key regulatory modules and highly connected hub genes [23, 24]. The above methods and their integration strategies have been successfully applied to the study of various complex traits in maize. For example, the combined analysis of dynamic transcriptomes and GWAS has revealed the regulatory mechanism of Agrobacterium-mediated transformation in maize [25]. In another case, the integration of QTL mapping and WGCNA uncovered the molecular mechanism by which *Zmccr3* regulates seed germination in maize [26]. In addition, the combined use of transcriptomics and WGCNA has provided strong support for the discovery and identification of elite genes. For example, this approach has been applied to the identification of genes responding to low temperature in maize seedlings [27]. It has also been used to study the plastic genetic regulation of root architecture in response to salt stress in maize [28]. Although these methods have been widely used in maize

research, their combined application to the regulation of husk number has not been reported yet.

Our integrated approach, two maize inbred lines with extreme differences in husk number were used. They were LJ311, with about 20 husks, and PHG35, with about 5 husks. A comprehensive transcriptome analysis was performed at four key developmental stages: V7, V8, V9, and V10. We adopted an integrated strategy that combined differential expression analysis, unsupervised hierarchical clustering, and WGCNA. Using this approach, we systematically identified co-expression modules and hub genes associated with husk number variation. By cross-validating unsupervised clustering with co-expression network analysis and applying a strict candidate gene filtering process, we constructed a bi-directional regulatory model. This model covers both promotive and inhibitory mechanisms and illustrates the molecular regulatory network controlling husk development. This study provides new insights into the molecular basis of husk number determination. It also offers valuable gene resources for breeding maize varieties suitable for mechanized harvesting.

Materials and methods

Plant Materials and Phenotypic Evaluation

Two maize inbred lines, LJ311 and PHG35, were used in this study. LJ311 has about 20 husk leaves, while PHG35 has about 5 husk leaves. They were grown in Tai'an, Shandong Province, China. Ear tissues including husks were collected at four developmental stages: V7, V8, V9, and V10. Three biological replicates were taken for each genotype at each stage. A total of 24 samples were obtained. All samples were immediately frozen in liquid nitrogen and stored at –80 °C.

RNA Extraction, Library Construction, and Sequencing

Total RNA extraction, quality control, and strand-specific cDNA library construction were performed by Novogene Co., Ltd. (Beijing, China). The qualified libraries were sequenced on the Illumina platform to generate 150 bp paired-end reads. All samples were processed in a single batch for library construction and sequencing. The raw sequencing data have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number PRJNA1491550.

Data Preprocessing and Identification of Husk-Specific Genes

Raw sequencing reads were first processed with fastp (v0.23.4) to remove low-quality reads and adapter sequences [29]. The cleaned reads were then aligned to the maize B73 reference genome (RefGen_v5) using HISAT2 (v2.2.1) [30]. Gene expression levels, including FPKM and count values, were quantified with StringTie (v2.1.7) [31].

Because the sampled tissues contained the entire ear, the initial expression matrix included transcripts from both husk and non-husk tissues. To enrich for genes specifically expressed in the husk, we compared all detected genes with a published spatiotemporal expression dataset of the maize ear [32]. Genes that overlapped with the ear-core dataset were removed. The remaining genes were considered as a candidate set of husk-expressed genes (Table S1). Further data cleaning was performed using the following criteria. First, the difference between the maximum and minimum FPKM values of a gene across all 24 samples had to be greater than 5 (min_max > 5). Second, within each biological group (genotype × stage combination, eight groups in total), the maximum coefficient of variation (CV = standard deviation / mean × 100%) was required to be less than 200% (max_cv < 200%). Third, each gene had to be reliably expressed in at least one biological group (reliable_groups > 0). After these filtering steps, a total of 12605 husk-specific genes were retained. This final gene set was used for all subsequent analyses, including differential expression analysis

and co-expression network construction.

Global Transcriptome Characterization and Differential Expression Analysis

To evaluate the overall relationships among the samples, we first defined a gene set that met a basic expression threshold. Genes with counts ≥ 10 in at least three samples were retained. These genes were used for principal component analysis (PCA) and Pearson correlation analysis. A PCA score plot and a clustered correlation heatmap were generated to visualize the sample groupings. In addition, library sizes and the numbers of detected genes were calculated for each sample. Bar plots for quality control and expression density curves were also drawn. The PCA, correlation analysis, and quality control plots described above were all based on the full gene set after basic filtering.

In contrast, the differential expression analysis and all subsequent functional analyses were performed using the filtered set of 12605 husk-specific genes. The raw counts of these 12605 genes were used as input for DESeq2 (v1.46.0). The design formula was set to ~ Stage + Genotype. Significantly differentially expressed genes (DEGs) were identified using the following criteria: |log2FoldChange| > 1 and *p*adj < 0.05. The DEGs were then divided into up-regulated and down-regulated groups based on the direction of their expression change. To verify the need for a stage-adjusted model, we also performed pairwise comparisons between LJ311 and PHG35 at each of the four developmental stages. The unique and shared DEGs across stages were visualized using an UpSet plot. Furthermore, a volcano plot, a pie chart showing DEG counts, and a heatmap of the top 100 DEGs were generated.

Unsupervised Hierarchical Clustering

To explore the natural grouping of highly variable genes, we selected the top 200 genes with

the largest variance across all 24 samples from the 12605 husk-specific genes. Each gene was then Z-score normalized. The distance between genes was calculated as 1 minus the Pearson correlation coefficient. Hierarchical clustering was performed using the complete-linkage method. Finally, the gene tree was cut into three clusters (Cluster 1–3). A clustered heatmap was generated using the pheatmap package (v1.0.12). The number of genes in each cluster was counted and displayed in a pie chart. To validate the late-stage upregulation pattern of Cluster 3, we selected the three genes with the largest |log2FoldChange| values from this cluster. Their expression levels across different genotypes and developmental stages were shown in a line plot. Error bars were used to represent the variation among the three biological replicates.

Weighted Gene Co-expression Network Analysis (WGCNA) and Cross-Validation

From the 12,605 husk-specific genes, we selected the top 8,000 genes with the largest variance for WGCNA analysis. The analysis was performed using the WGCNA package (v1.72-1). Network topology was evaluated with the pickSoftThreshold function. Based on the scale-free topology fit index ($R^2 > 0.85$) and the mean connectivity, we chose a soft-thresholding power of 10. A signed network was then constructed using the blockwiseModules function with the following parameters: minModuleSize = 30 and mergeCutHeight = 0.25. Module eigengenes were calculated for each module. Pearson correlation analysis was then performed between the module eigengenes and two traits: husk number and genotype.

Based on the correlation results, we identified three modules that were most significantly associated with husk number. The blue and brown modules were positively correlated, whereas the turquoise module was negatively correlated with this trait. A module–trait correlation heatmap and line plots showing the expression trends of these module eigengenes across developmental stages

were generated. GO and KEGG enrichment analyses were performed for all genes in the three key modules using the Omicshare online tool (www.omicshare.com/tools). All analyses were run with the default parameters. In addition, the module membership (kME) and gene significance (GS) were calculated for each gene. Strict hub genes were defined as those satisfying both $|kME| > 0.9$ and $|GS| > 0.7$. Scatter plots of kME versus GS were then generated for each module, with the hub genes highlighted.

To cross-validate the results of unsupervised clustering and WGCNA, we took the 44 genes from Cluster 3 and intersected them with the gene lists of the blue and brown modules, separately. The overlaps were visualized using an UpSet plot. In addition, all 44 genes of Cluster 3 were independently subjected to GO and KEGG enrichment analysis.

Screening, Expression Visualization, and Validation of High-Confidence Candidate Genes

The hub genes from the positively correlated modules (blue and brown) were first combined into one set. This combined set was then intersected with the up-regulated DEGs that had $|log2FoldChange| > 2$. This intersection gave us the up-regulated high-confidence candidate gene set. Similarly, the hub genes from the turquoise module were intersected with the down-regulated DEGs having $|log2FoldChange| > 2$. This produced the down-regulated candidate gene set. A stricter fold-change cutoff was used here to focus on the genes with the largest expression differences. The entire screening process was illustrated using a Venn diagram. GO and KEGG enrichment analyses were performed separately for the up-regulated and down-regulated candidate gene sets. The results were visualized using bubble plots to show the functional differences between the two groups.

From the final candidate gene sets, we selected the top 30 genes with the largest $|log2FoldChange|$ values from each direction. A heatmap was generated to display their expression

patterns. In addition, three of the most significant hub genes were chosen from each of the up-regulated and down-regulated sets. Their expression levels across stages and genotypes were shown in line plots. Error bars in these plots represent the variation among the biological replicates. Using the igraph (v1.4.3) and ggraph (v2.1.0) packages, we extracted the top 10 hub genes with the highest |kME| values from the blue, brown, and turquoise modules. Their first-order neighbors (Pearson correlation coefficient > 0.5) were also included. Co-expression subnetworks were then constructed for each of the three modules.

To validate the reliability of the RNA-seq data, five key genes were selected from the final candidate gene list. Primers for these genes were designed using the qPrimerdb database (https://qprimerdb.biodb.org/) and are listed in Table S5. The maize *Cullin* gene was used as the internal reference. Quantitative real-time PCR was performed using the SYBR Green method. Relative expression levels were calculated using the $2^{-\Delta\Delta Ct}$ method.

Results

Phenotypic Difference in Husk Number and Global Transcriptome Characteristics

The two inbred lines showed a significant difference in mature husk number. LJ311 had an average of about 20 husks, while PHG35 had about 5 husks (Figure 1A, B). After data filtering, a total of 12605 husk-specific genes were retained for subsequent differential expression and co-expression network analyses. To evaluate the overall relationships among the 24 samples, we performed principal component analysis (PCA) and Pearson correlation analysis. For these analyses, we used all genes that met the basic expression threshold. A gene was considered expressed if it had counts ≥ 10 in at least three samples.

The PCA results showed a clear separation by genotype along the first principal component (PC1). PC1 explained 48.3% of the total variance. The second principal component (PC2) explained 12.3% of the variance and further reflected the effect of developmental stage (Figure 1C). The three biological replicates of the same genotype-stage combination clustered tightly together on the PCA plot. Pearson correlation analysis further supported this pattern. Within each biological group, the correlation coefficients among replicates were all above 0.95. Meanwhile, the correlations between samples of different genotypes were obviously lower (Figure 1D). The library size of each sample ranged from 27 to 40 million reads. The number of detected genes was close to 30000 in all samples. In addition, the expression density curves of the 24 samples were highly overlapping (Figure S1). Detailed information on RNA-seq data quality is provided in Table S2. Together, these results indicate that the transcriptome data were of high quality and suitable for the following quantitative analyses.

Identification of Differentially Expressed Genes Between LJ311 and PHG35

A stage-adjusted DESeq2 model was used to identify differentially expressed genes (DEGs) between LJ311 and PHG35. The design formula was set to ~ Stage + Genotype. Genes with |log2FoldChange| > 1 and padj < 0.05 were considered significantly differentially expressed. Using these criteria, a total of 3764 DEGs were identified. Among them, 1900 genes were up-regulated in LJ311, while 1864 genes were down-regulated (Figure 2A, B). The numbers of up- and down-regulated DEGs were nearly equal. For most DEGs, the |log2FC| values fell within the range of 1 to 10. The proportions of up- and down-regulated genes were roughly balanced. Stage-specific comparisons further showed that there were both shared and unique DEGs across the four developmental stages (Figure S4). This observation supported the use of the stage-adjusted model

to capture the robust genotype-driven differences. A heatmap of the top 100 most significant DEGs clearly separated the LJ311 and PHG35 samples. The up-regulated and down-regulated genes each formed their own clusters (Figure 2C). These results indicate that extensive transcriptional reprogramming occurred between the two lines. The complete list of DEGs is provided in Table S3.

Unsupervised Hierarchical Clustering Identified a Late-Stage-Specific Gene Cluster

To explore the natural grouping of highly variable genes, we performed hierarchical clustering on the top 200 genes with the largest variance across all 24 samples. These genes were divided into three clusters (Cluster 1–3) (Figure 3A). Cluster 1 contained 80 genes and was mainly up-regulated in PHG35. Cluster 2 contained 76 genes and showed relatively high expression across all stages in LJ311. By comparison, Cluster 3 (44 genes) showed a unique expression pattern. It was strongly and specifically up-regulated at the late developmental stages (V8–V10) in LJ311. Meanwhile, its expression remained very low in PHG35 and also at the V7 stage in LJ311 (Figure 3A, B). The three genes with the largest |log2FC| values in Cluster 3 all showed a clear increase in expression starting from V8 in LJ311. However, no such increase was found in PHG35 (Figure 3C). This late-stage-specific activation suggests that Cluster 3 may represent a functionally coordinated gene set. It is likely related to the determination of husk number.

WGCNA Identified Co-expression Modules Significantly Correlated with Husk Number

WGCNA was performed using the top 8000 highly variable genes. A soft-thresholding power of 10 was chosen. At this power, the scale-free topology fit index (R^2) exceeded 0.85, and the mean connectivity reached a plateau (Figure 4A, B). In total, ten co-expression modules were obtained. Pearson correlation analysis between module eigengenes (MEs) and traits identified three modules

that were significantly correlated with husk number. The blue module (1371 genes, $r = 0.88$, $p < 0.001$) and the brown module (1183 genes, $r = 0.85$, $p < 0.001$) were positively correlated with both husk number and the LJ311 genotype. In contrast, the turquoise module (1963 genes, $r = -0.97$, $p < 0.001$) was strongly negatively correlated with this trait (Figure 4C). The ME expression trends of these three modules confirmed the correlation results. The blue and brown modules were consistently expressed at higher levels in LJ311, and the blue module showed a further increase at the V8–V10 stages. By comparison, the turquoise module was predominantly expressed in PHG35 (Figure 4D). The ME heatmap of all modules together with the gene dendrogram and module color assignment further illustrated the overall modular structure of the co-expression network (Figure S2).

Cross-Validation Between Hierarchical Clustering and WGCNA

To evaluate the consistency between the unsupervised clustering and the co-expression network, we intersected the 44 genes of Cluster 3 with the gene lists of the blue and brown modules. Cluster 3 shared only seven genes with the blue module and had no overlap with the brown module (Figure 5A). In comparison, out of the 80 genes in Cluster 1, 79 belonged to the turquoise module. The genes in Cluster 2 were mainly found in the blue and brown modules (Figure S3). Taken together, these overlaps indicate that the two methods showed an overall consistency. Although the overlap was limited, GO and KEGG enrichment analyses of all 44 genes in Cluster 3 showed that they were significantly enriched in several pathways. These pathways included the TCA cycle, oxidative phosphorylation, DNA replication and repair, glycerophospholipid metabolism, and carotenoid/ABA metabolism (Figure 5B). Together, these functional categories point to energy production, cell proliferation, and membrane biosynthesis. This suggests that Cluster 3 acts as a

late-stage functional module that promotes husk growth.

Hub Gene Identification and Refinement of High-Confidence Candidate Genes

Genes satisfying both |kME| > 0.9 and |GS| > 0.7 were defined as strict hub genes. In the blue, brown, and turquoise modules, 123, 231, and 282 strict hub genes were identified, respectively (Figure 6A). These hub genes are the most highly connected core nodes within each module. They also have the strongest correlations with husk number. To further verify the central roles of these hub genes within their modules, we constructed co-expression subnetworks. Each subnetwork was built using the top 10 hub genes of a module and their first-order neighbors. The results further confirmed that these hub genes occupy core positions in their respective modules (Figure S5).

The hub genes from the positively correlated modules (blue and brown) were first merged into one set. This set was then intersected with the up-regulated DEGs having |log2FC| > 2. This gave 324 up-regulated high-confidence candidate genes (Figure 6B). Similarly, the hub genes from the turquoise module were intersected with the down-regulated DEGs having |log2FC| > 2. This produced 266 down-regulated high-confidence candidate genes (Figure 6E). The complete list of final candidate genes is provided in Table S4.

GO and KEGG enrichment analyses revealed clear functional differences between the two groups of candidate genes. The up-regulated candidates were significantly enriched in several pathways. These pathways included DNA replication, mismatch repair, homologous recombination, brassinosteroid (BR) biosynthesis, MAPK signaling pathway-plant, ribosome, and glycerophospholipid metabolism. These pathways are mainly related to cell cycle progression, hormone signaling, and membrane biosynthesis (Figure 6C, D). In contrast, the down-regulated candidates were enriched in pathways such as glycolysis, fructose and mannose metabolism, xylan

degradation, MAPK signaling, and phosphatidylinositol signaling system (Figure 6F, G). These results indicate that the increase in husk number in LJ311 was driven by cell proliferation and the activation of growth-promoting pathways. In PHG35, however, energy metabolism was shifted, and growth-inhibitory signals were strengthened.

Expression Patterns and Preliminary Validation of Final Candidate Genes

To visualize the overall expression landscape of the high-confidence candidate genes, a heatmap was created. We selected the top 30 genes with the largest |log2FC| values from each of the up- and down-regulated directions. The heatmap clearly divided the candidate genes into two major clusters. This separation perfectly matched their regulatory directions (Figure 7A). Within each cluster, the expression patterns were highly consistent with the sample groupings. The up-regulated candidate genes showed relatively high expression across all developmental stages in LJ311. In contrast, the down-regulated candidates displayed the opposite trend and were mainly expressed at higher levels in PHG35. This sharp contrast further confirmed the reliability of the final candidate gene set and its strong association with the husk number phenotype.

To further verify the reliability of the RNA-seq results, we analyzed the fold changes of five representative high-confidence candidate genes using both RNA-seq and qRT-PCR. The results showed that the expression trends obtained by qRT-PCR were highly consistent with those from RNA-seq. The fold change values from the two methods matched well across all four developmental stages (Figure 7B). This independent validation increased the confidence in the candidate genes identified in this study. It also suggests that these genes may be the true regulators of husk number.

Discussion

The Advantage of Multi-Method Integration in Mining Genes for Complex Traits

Husk number in maize is a complex quantitative trait controlled by many genes and multiple pathways. A single method often fails to capture the complete regulatory network. In this study, we integrated three approaches—differential expression analysis, unsupervised hierarchical clustering, and WGCNA—to improve the efficiency of candidate gene discovery.

The order of integration has a significant impact on the results. Studies have shown that first building the co-expression network and then adding the differential expression information preserves the global topological structure of the network better. This helps reduce false positives and captures more biologically relevant modules and hub genes [33, 34]. The results of this study also support this idea. Cluster 3, found by hierarchical clustering, shared only seven genes with the blue module. However, their functional enrichment was highly consistent. This indicates that the unsupervised method successfully detected a functionally consistent subset within a large co-expression module. Such multi-strategy integration has been verified in several plant systems. For example, it was used to screen key genes for salt stress response in rice [33], to identify modules for cold tolerance in maize seedlings [27], and to recognize functional modules under abiotic stress in poplar leaves [35]. These examples demonstrate the robustness of this approach across different research backgrounds.

Functional Significance of the Late-Stage-Specific Gene Cluster, Cluster 3

Cluster 3 contained 44 genes and was specifically up-regulated at the V8–V10 stages in LJ311. This is one of the core findings of this study. This period coincides exactly with the key stage of rapid differentiation and elongation of husk primordia. Functional enrichment analysis showed that

Cluster 3 was significantly enriched in the TCA cycle, oxidative phosphorylation, DNA replication, glycerophospholipid metabolism, and carotenoid/ABA metabolism. In maize, TCA cycle genes show tissue-specific expression patterns. Their enzyme activities are essential for supplying energy and carbon skeletons during late development [36, 37]. The TCA cycle is coupled with oxidative phosphorylation. Together, they provide ATP and reducing power for biosynthesis. Additionally, the TCA cycle is more than just an energy provider. Its disruption can trigger extensive metabolic reprogramming and activate stress responses. This has been shown in mammalian systems [38]. In plants, the TCA cycle also plays non-canonical roles in developmental transitions [39]. The enrichment of DNA replication genes suggests that active cell proliferation or endoreduplication still occurs in late-stage husk tissues. This is very important for organ morphogenesis [40]. Glycerophospholipid metabolism provides basic components for membrane structures. Its intermediate products, such as phosphatidic acid, can also act as signaling molecules. They take part in membrane biogenesis and signal transduction [41-43]. Together, these pathways form a coordinated molecular network. They support the late development of the husk in several ways, from energy supply and cell proliferation to membrane biosynthesis.

Functional Divergence of Positively and Negatively Regulated Modules

The three WGCNA modules that were significantly correlated with husk number showed distinct regulatory functions. The blue module was enriched in cell cycle, DNA replication, and protein homeostasis. The brown module was enriched in actin filament organization, Golgi organization, and brassinosteroid (BR) biosynthesis. Although both modules were positively correlated with husk number, their functional divergence suggests that they are involved in different biological processes. Positive regulation may require the coordination of the cell proliferation

machinery and the cytoskeleton–hormone signaling system.

BR is an important hormone that regulates plant growth and development. There is cross-talk between BR and the MAPK pathway. For example, in stomatal development, BR signaling inhibits the MAPK pathway through GSK3 kinase [44]. In maize, the transcription factor *ZmRAVL1* can directly activate the expression of BR biosynthesis genes. These genes include *lac1*, *ZmD11*, and *brd1*. At the same time, phytochrome phyA senses the ratio of red to far-red light. This light signal then regulates the protein stability of *ZmRAVL1*. In this way, environmental light signals are coupled with internal BR levels. This allows a dynamic regulation of canopy architecture [45, 46]. In contrast, the turquoise module was mainly expressed in the low-husk line PHG35. This module was enriched in endoplasmic reticulum (ER) stress and protein phosphorylation. The ER can activate BR signaling through the unfolded protein response (UPR). In this way, stress signals and developmental signals are integrated [47]. The enrichment of ER stress in the negative module suggests something important. Rapid husk growth may create a high demand for protein synthesis. This pressure could activate the ER quality control system. As a result, organ development may be limited to some degree. This antagonistic module architecture indicates that husk number is controlled by a bi-directional regulatory network. This network involves BR/MAPK cross-talk, the integration of ER stress, and epigenetic regulation.

High-Confidence Candidate Genes Define the Molecular Boundary Between Promotion and Inhibition of Husk Number

By integrating module hub genes with high-magnitude differentially expressed genes, we obtained a compact set of high-confidence candidate genes. The up-regulated candidates were mainly enriched in DNA replication, BR biosynthesis, and glycerophospholipid metabolism. This

is consistent with the overall functional features of the positively correlated modules described earlier. Among the up-regulated candidates, several genes encode enzymes and transcription factors. These genes may act downstream of the maize BR signaling cascade. BR has been shown to regulate plant development by promoting cell cycle progression [48, 49]. On the other hand, the down-regulated candidates were enriched in glycolysis, xylan degradation, and MAPK signaling. The enrichment of both an energy metabolism pathway and a cell wall remodeling process at the same time suggests something important. It suggests that a state of metabolic inhibition may exist in the low-husk line. In this state, basic energy production is reduced. Cell wall loosening is also restricted. As a result, the initiation of husk primordia is suppressed. In maize, glycoside hydrolases such as MLGH1 take part in the degradation and remodeling of cell wall polysaccharides [50, 51]. The specific enrichment of MAPK signaling among the down-regulated candidates further supports the bi-directional regulatory model. This is consistent with the dual functions of plant MAPK cascades. They can both promote and restrict growth [52]. This antagonistic molecular dialogue was also reflected at the transcriptional level. It is likely that that transcription factors may play important roles in the bi-directional regulation of husk number. Taken together, the high-confidence candidate genes clearly define the molecular boundary between “promotion” and “inhibition” in husk number regulation. Upstream BR and MAPK signals converge on these executive-level genes. They coordinate DNA replication, energy metabolism, and cell wall dynamics. In the end, these genes determine the phenotypic output of husk number.

Comparison with Known Regulatory Mechanisms of Husk Development and the Incremental Contribution of This Study

In recent years, association-based studies have revealed the genetic basis of variation in husk

traits. For example, the *ZMET2* gene encodes a DNA methyltransferase. A 10-bp insertion/deletion polymorphism in its 3'UTR region can change the transcription level. This in turn regulates the number of husk leaves [53]. The *RHW1-ZCN4* pathway regulates husk width through a MYB-like transcriptional repressor [54]. The *ZmNRT2.5* gene encodes a nitrate transporter. Its variation is related to both husk width and the efficiency of nitrate transport to developing grains [55]. Some progress has been made in understanding the genetic regulation of husk development in maize. The gene *tasselsheath4* (*tsh4*) encodes an SBP-box transcription factor. It is a core regulator that inhibits husk growth and promotes inflorescence branching [56]. This gene forms a hierarchical regulatory network together with *tsh1* and *lg2*. In this network, *tsh4* directly binds to and activates the expression of downstream genes. Together, they control the inhibition of husk growth and the formation of branches [57]. These studies have begun to reveal the transcriptional, epigenetic, and metabolic regulatory networks involved in husk development.

Notably, the classical meristem identity genes, such as *tsh4*, were not directly detected in our final candidate gene set. This suggests that there may be another mechanism for husk number variation. This mechanism may be independent of the early meristem identity decisions. The candidate genes we identified focused more on late developmental events. These events include cell proliferation, energy metabolism, and the integration of hormone signals. These genes may act downstream of the early-acting genes like *tsh4*. For example, the BR signaling components enriched in the up-regulated candidates may receive signals from the early meristem identity program. Then they promote cell cycle progression. This drives the continuous proliferation of husk primordia. On the other hand, the MAPK signaling and cell wall degradation genes enriched in the down-regulated candidates may act as negative regulatory effectors. They may restrict the overgrowth of husk leaves.

This model of "early decision – late execution" integrates the known regulatory factors of husk development with our findings. It provides a more complete view for understanding the molecular mechanisms of husk number variation.

Conclusions

This study systematically explored the molecular basis of husk number variation in maize by integrating differential expression analysis, unsupervised hierarchical clustering, and WGCNA. Three co-expression modules (blue, brown, and turquoise) were identified that significantly correlated with husk number and represented growth-promoting and growth-inhibiting networks, respectively. Hierarchical clustering revealed a small, distinct gene cluster (Cluster 3) that was strongly activated at late developmental stages (V8–V10) in the high-husk line LJ311 and enriched in the TCA cycle, oxidative phosphorylation, DNA replication, and membrane lipid metabolism. By intersecting module hub genes with high-magnitude differentially expressed genes, we obtained a set of high-confidence candidate genes. The up-regulated candidates were enriched in cell cycle, brassinosteroid (BR) signaling, and phospholipid metabolism, whereas the down-regulated candidates were enriched in MAPK signaling, glycolysis, and cell wall polysaccharide degradation. Taken together, these results suggest that husk number is controlled by a bidirectional regulatory program that converges on a BR–MAPK–cell cycle axis. The candidate genes identified in this work provide valuable targets for future functional validation and molecular breeding. Further studies should focus on the functional characterization of these hub genes and their upstream regulators.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article and its additional files. The raw RNA-seq data are available in the NCBI BioProject repository, accession number PRJNA1491550 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1491550).

Competing interests

The authors declare that they have no competing interests

Acknowledgments

Not applicable.

Funding

This work was supported by the National Natural Science Foundation of China to Cuixia Chen (No. 32370672), the National Key R&D Program of China to Cuixia Chen (2023YFD1200501).

Authors' Contributions

XX and CC conceived and designed the experiments. XX performed the experiments, analyzed the data, and drafted the manuscript. LD and WY performed phenotyping. SH and SL cultivated the plant materials. XX and HY collected plant samples. All authors participated in manuscript revision and approved the final version.

Supplementary Information

Additional supporting information may be found in the online version of this article.

Additional file 1: Table S1. Filtered gene expression matrix of maize husk samples.

Additional file 2: Table S5. Sequence of primers used in qRT-PCR.

Additional file 3: Figure S1. RNA-seq quality control metrics of all genes passing the basic expression threshold. (A) Bar plot showing the library size (in million reads) for each of the 24 samples. (B) Bar plot showing the number of detected genes per sample. (C) Density plot of log2(counts + 1) expression values across all 24 samples.

Additional file 4: Table S2. Summary of RNA-seq read statistics for all 24 samples.

Additional file 5: Figure S4. Stage-specific DEGs identified by pairwise comparisons at each developmental stage(V7: green, V8: orange, V9: purple, V10: pink). The UpSet plot displays the number of unique and shared DEGs across the four stages.

Additional file 6: Table S3. Complete list of differentially expressed genes (DEGs) between LJ311 and PHG35.

Additional file 7: Figure S2. Overview of all WGCNA modules. (A) Heatmap of module eigengene (ME) expression values across all 24 samples. Each row represents one co-expression module, and

each column represents one sample. Red indicates high expression, and blue indicates low expression. (B) Complete gene hierarchical clustering dendrogram with module color assignment. Each branch tip represents an individual gene, and the color bar below indicates its module membership.

Additional file 8: Figure S3. UpSet plots of overlaps between Cluster 1/Cluster 2 and WGCNA modules. (A) Overlap between Cluster 1 (green) and the turquoise module (turquoise). (B) Overlap between Cluster 2 (light blue) and the blue module (blue) and the brown module (brown).

Additional file 9: Figure S5. Co-expression subnetworks of the top 10 hub genes in the three key modules. (A) Blue module. (B) Brown module. (C) Turquoise module. All ten hub genes are displayed as colored nodes, with node colors corresponding to the module identity (blue, brown, and turquoise, respectively). Node size is proportional to the absolute module membership (|kME|). Gray edges represent Pearson correlations between connected genes ($|r| > 0.5$), with edge width indicating the strength of the correlation. Only the top three hub genes with the highest |kME| values in each module are labeled.

Additional file 10: Table S4. Final candidate genes.

References

1. Li S, Wang K, Xie R, Ming B. Mechanical grain harvesting promotes the transformation of maize production mode [in Chinese]. Sci Agric Sin. 2018;51:1842–1844.

2. Xie R, Ming B, Shang G, Wang K, Hou P, Li, S. Current state and suggestions for mechanical harvesting of corn in China. J Integr Agric. 2022;21:892–897.

3. Liu F, Wang K, Li J, Wang X, Sun Y, Chen Y, et al. Analysis of factors affecting the quality of

509 mechanical grain harvesting in maize [in Chinese]. Crops. 2013;29:116–119.

510 4. Liu J, Yu H, Liu Y, Deng S, Liu Q, Liu B, et al. Genetic dissection of grain water content and
511 dehydration rate related to mechanical harvest in maize. BMC Plant Biol. 2020;20:118.

512 5. Dai L, Wu L, Dong Q, Zhang Z, Wu N, Song Y, et al. Genome-wide association study of field
513 grain drying rate after physiological maturity based on a resequencing approach in elite maize
514 germplasm. Euphytica. 2017;213:182.

515 6. Magarini A, Colombo F, Cassani E, Ghidoli M, Landoni M, Sangiorgio S, et al. The role of
516 husk traits in maize susceptibility to Fusarium verticillioides: A multi-location study in northern
517 Italy. Food Energy Sec. 2024;13:e537.

518 7. Sala RG, Andrade FH, Camadro EL, Cerono JC. Quantitative trait loci for grain moisture at
519 harvest and field grain drying rate in maize (*Zea mays* L.). Theor Appl Genet. 2006;112:462–471.

520 8. Yu Y, Li W, Liu Y, Liu Y, Zhang Q, Ouyang Y, et al. A Zea genus-specific micropeptide controls
521 kernel dehydration in maize. Cell. 2025;188:44–59.e21.

522 9. Chuck G, Whipple C, Jackson D, Hake S. The maize SBP-box transcription factor encoded by
523 tasselsheath4 regulates bract development and the establishment of meristem boundaries.
524 Development. 2010;137:1243–1250.

525 10. Lejeune P, Bernier G. Effect of environment on the early steps of ear initiation in maize (*Zea*
526 *mays* L.). Plant Cell Environ. 1996;19:217–224.

527 11. Ji HC, Brewbaker JL. Inheritance of long husk leaves in maize. Maydica. 2007;52:101–108.

528 12. Shi W, Shao H, Sha Y, Shi R, Shi D, Chen Y, et al. Grain dehydration rate is related to post-
529 silking thermal time and ear characters in different maize hybrids. J Integr Agric. 2022;21:964–976.

530 13. Troyer AF, Ambrose WB. Plant Characteristics Affecting Field Drying Rate of Ear Corn. Crop

Science. 1971;11:529-531.

14. Li L, Ming B, Xie R , Wang K , Hou P, Li S. Differences in ear traits of maize varieties and their effects on grain dehydration [in Chinese]. Sci Agric Sin. 2018;51:1855–1867.

15. Wang K, Li S. Analysis of factors affecting grain dehydration rate in maize [in Chinese]. Sci Agric Sin. 2017;50:2027–2035.

16. Cui Z, Xia A, Zhang A, Luo J, Yang X, Zhang L, et al. Linkage mapping combined with association analysis reveals QTL and candidate genes for three husk traits in maize. Theor Appl Genet. 2018;131:2131–2144.

17. Liang Z, Xi N, Liu T, Li M, Sang M, Zou C, et al. A combination of QTL mapping and genome-wide association study revealed the key gene for husk number in maize. Theor Appl Genet. 2024;137:112.

18. Zhang X, Lu M, Xia A, Xu T, Cui Z, Zhang R, et al. Genetic analysis of three maize husk traits by QTL mapping in a maize-teosinte population. BMC Genomics. 2021;22:386.

19. Zhou G, Mao Y, Xue L, Chen G, Lu H, Shi M, et al. Genetic dissection of husk number and length across multiple environments and fine-mapping of a major-effect QTL for husk number in maize (*Zea mays* L.). Crop J. 2020;8:1071–1080.

20. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009;10:57–63.

21. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci. 1998;95:14863–14868.

553 23. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis.
554 BMC Bioinf. 2008;9:559.
555 24. Zhang B, Horvath S. A General Framework For Weighted Gene Co-Expression Network
556 Analysis. Stat Appl Genet Mol Biol. 2005;4:1–13.
557 25. Liu M, Yang Y, Liang T, Hou F, Zhang M, He S, et al. Dynamic transcriptome and GWAS
558 uncover a hydroxyproline-rich glycoprotein that suppresses Agrobacterium-mediated
559 transformation in maize. Mol Plant. 2025;18:747–764.
560 26. Feng L, Zhou M, Tao A, Ma X, Wang N, Zhang H, et al. Map-based cloning of Zmccr3 and its
561 network construction and validation for regulating maize seed germination. Theor Appl Genet.
562 2025;138:105.
563 27. Yu T, Zhang J, Cao J, Ma X, Li W, Yang G. Hub Gene Mining and Co-Expression Network
564 Construction of Low-Temperature Response in Maize of Seedling by WGCNA. Genes.
565 2023;14:1598.
566 28. Li P, Yang X, Wang H, Pan T, Wang Y, Xu Y, et al. Genetic control of root plasticity in response
567 to salt stress in maize. Theor Appl Genet. 2021;134:1475–1492.
568 29. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
569 Bioinformatics. 2018;34:i884–i890.
570 30. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of
571 RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11:1650–1667.
572 31. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables
573 improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–
574 295.

575 32. Shen X, Xiao B, Kaderbek T, Lin Z, Tan K, Wu Q, et al. Dynamic transcriptome landscape of
576 developing maize ear. Plant J. 2023;116:1856–1870.
577 33. Riccio-Rengifo C, Finke J, Rocha C. Identifying stress responsive genes using overlapping
578 communities in co-expression networks. BMC Bioinf. 2021;22:541.
579 34. Sánchez-Baizán N, Ribas L, Piferrer F. Improved biomarker discovery through a plot twist in
580 transcriptomic data analysis. BMC Biology. 2022;20:208.
581 35. Zhang H, Yin T. Identifying hub genes and key functional modules in leaf tissue of Populus
582 species based on WGCNA. Genetica. 2024;153:5.
583 36. Liu Y, Qu J, Zhang L, Xu X, Wei G, Zhao Z, et al. Identification and characterization of the
584 TCA cycle genes in maize. BMC Plant Biol. 2019;19:592.
585 37. Zhang T, Noll SE, Peng JT, Klair A, Tripka A, Stutzman N, et al. Chemical imaging reveals
586 diverse functions of tricarboxylic acid metabolites in root growth and development. Nat Commun.
587 2023;14:2567.
588 38. Ryan DG, Yang M, Prag HA, Blanco GR, Nikitopoulou E, Segarra-Mondejar M, et al.
589 Disruption of the TCA cycle reveals an ATF4-dependent integration of redox and amino acid
590 metabolism. eLife. 2021;10:e72593.
591 39. Zhang T, Peng JT, Klair A, Dickinson AJ. Non-canonical and developmental roles of the TCA
592 cycle in plants. Curr Opin Plant Biol. 2023;74:102382.
593 40. De Veylder L, Larkin JC, Schnittger A. Molecular control and function of endoreplication in
594 development and physiology. Trends Plant Sci. 2011;16:624–634.
595 41. Nakamura Y. Plant Phospholipid Diversity: Emerging Functions in Metabolism and Protein–
596 Lipid Interactions. Trends Plant Sci. 2017;22:1027–1040.

42. Xiao Q, Singer SD, Nakamura Y, Chen G. Phospholipids in plant systems: metabolism, regulation and functional insights. Prog Lipid Res. 2026;101:101377.

43. Zhang W, Qin C, Zhao J, Wang X. Phospholipase D alpha 1-derived phosphatidic acid interacts with ABI1 phosphatase 2C and regulates abscisic acid signaling. Proc Natl Acad Sci. 2004;101:9508–9513.

44. Kim T-W, Michniewicz M, Bergmann DC, Wang Z-Y. Brassinosteroid regulates stomatal development by GSK3-mediated inhibition of a MAPK pathway. Nature. 2012;482:419–422.

45. Tian J, Wang C, Chen F, Qin W, Yang H, Zhao S, et al. Maize smart-canopy architecture enhances yield at high densities. Nature. 2024;632:576–584.

46. Tian J, Wang C, Xia J, Wu L, Xu G, Wu W, et al. Teosinte ligule allele narrows plant architecture and enhances high-density maize yields. Science. 2019;365:658–664.

47. Che P, Bussell JD, Zhou W, Estavillo GM, Pogson BJ, Smith SM. Signaling from the endoplasmic reticulum activates brassinosteroid signaling and promotes acclimation to stress in Arabidopsis. Sci Signal. 2010;141: ra69.

48. Nolan TM, Vukašinović N, Liu D, Russinova E, Yin Y. Brassinosteroids: Multidimensional Regulators of Plant Growth, Development, and Stress Responses. Plant cell. 2020;32:295–318.

49. Planas-Riverola A, Gupta A, Betegón-Putze I, Bosch N, Ibañes M, Caño-Delgado AI. Brassinosteroid signaling in plant development and adaptation to stress. Development. 2019;146:dev151894.

50. Kraemer FJ, Lunde C, Koch M, Kuhn BM, Ruehl C, Brown PJ, et al. A mixed-linkage (1,3;1,4)-β-D-glucan specific hydrolase mediates dark-triggered degradation of this plant cell wall polysaccharide. Plant Physiol. 2021;185:1559–1573.

619 51. Nazipova A, Gorshkov O, Eneyskaya E, Petrova N, Kulminskaya A, Gorshkova T, et al.
620 Forgotten Actors: Glycoside Hydrolases During Elongation Growth of Maize Primary Root. Front.
621 Plant Sci. 2022;Volume 12–2021.
622 52. Komis G, Šamajová O, Ovečka M, Šamaj J. Cell and Developmental Biology of Plant
623 Mitogen-Activated Protein Kinases. Annu Rev Plant Biol. 2018;69:237–265.
624 53. Wang Z, Xia A, Wang Q, Cui Z, Lu M, Ye Y, et al. Natural polymorphisms in ZMET2 encoding
625 a DNA methyltransferase modulate the number of husk layers in maize. Plant Physiol.
626 2024;195:2129–2142.
627 54. Xia A, Zheng L, Wang Z, Wang Q, Lu M, Cui Z, et al. The RHW1-ZCN4 regulatory pathway
628 confers natural variation of husk leaf width in maize. New Phytol. 2023;239:2367–2381.
629 55. Wang Q, Wang M, Xia A, Wang J, Wang Z, Xu T, et al. Natural variation in ZmNRT2.5
630 modulates husk leaf width and promotes seed protein content in maize. Plant Biotechnol. J. 2025;23
631 4:1039–1052.
632 56. Xiao Y, Guo J, Dong Z, Richardson A, Patterson E, Mangrum S, et al. Boundary domain genes
633 were recruited to suppress bract growth and promote branching in maize. Sci Adv. 2022;8:eabm6835.
634 57. Chuck G, Bortiri E. The unique relationship between tsh4 and ra2 in patterning floral
635 phytomers. Plant Signal Behav. 2010;5:979–981.

636 ## Figure legends

637 **Figure 1. Phenotypic difference and global transcriptomic characterization of husk number.**
638 (A) Mature husk phenotypes of the two parental lines. Left, PHG35; right, LJ311. (B) Mean husk
639 number of PHG35 and LJ311 across multiple years and locations. (C) PCA of the 24 samples based
640 on all genes passing the basic expression threshold. Colors indicate genotypes (LJ311, red; PHG35,

blue), and shapes indicate developmental stages (V7, circle; V8, triangle; V9, diamond; V10, square). (D) Pearson correlation heatmap with hierarchical clustering of the 24 samples. Darker blue indicates higher correlation.

Figure 2. Identification and expression profiling of differentially expressed genes (DEGs) between LJ311 and PHG35. (A) Volcano plot of DEGs. The screening thresholds were set at $|\log2FC| > 1$ and $padj < 0.05$. Red, up-regulated; blue, down-regulated; gray, not significant. (B) Pie chart showing the proportion of up-regulated and down-regulated DEGs. (C) Heatmap of the top 100 DEGs. Red indicates high expression, and blue indicates low expression.

Figure 3. Unsupervised hierarchical clustering analysis of highly variable genes. (A) Heatmap of the top 200 most variable genes across all 24 samples. Genes were Z-score normalized and clustered into three major groups (Cluster 1–3) based on their expression patterns. (B) Pie chart showing the number of genes in each of the three clusters. (C) Expression dynamics of three representative genes from Cluster 3 across developmental stages and genotypes. These three genes had the largest |log2FC| values within the cluster. Data are presented as mean ± SE.

Figure 4. WGCNA of husk number-related modules. (A, B) Scale independence (A) and mean connectivity (B) across soft-thresholding powers. The red line in (A) marks $R^2 = 0.85$. (C) Module–trait correlation heatmap. Blue, brown, and turquoise modules were most significantly associated with husk number. *** $p < 0.001$. (D) Expression trends of module eigengenes for the three modules across developmental stages. Red, LJ311; blue, PHG35. Data are mean ± SE.

Figure 5. Cross-validation between hierarchical clustering and WGCNA. (A) UpSet plot showing the overlap of genes between Cluster 3 and the blue and brown modules. (B) GO enrichment bubble plot of all 44 genes in Cluster 3. Bubble size represents the number of enriched

genes, and bubble color represents the adjusted *p*-value (p_{adj}). (C) KEGG enrichment bubble plot of all 44 genes in Cluster 3. Bubble size represents the number of enriched genes, and bubble color represents the adjusted *p*-value (p_{adj}).

Figure 6. Hub gene identification and functional enrichment of high-confidence candidates. (A) kME–GS scatter plots for the blue, brown, and turquoise modules. Dashed lines mark |kME| = 0.9 and |GS| = 0.7; red dots, hub genes. (B, E) Venn diagrams showing intersections between hub genes and DEGs (|log2FC| > 2) for positively (B) and negatively (E) correlated modules. (C, D) GO (C) and KEGG (D) enrichment bubble plots of up-regulated candidates. (F, G) GO (F) and KEGG (G) enrichment bubble plots of down-regulated candidates. For all bubble plots, bubble size indicates gene count, and color indicates adjusted *p*-value (p_{adj}).

Figure 7. Expression patterns and experimental validation of final candidate genes. (A) Heatmap of the top 30 up- and down-regulated candidate genes. Genes are clustered into two groups based on their regulatory directions. (B) Fold change comparison between RNA-seq and qRT-PCR for five representative hub genes across four developmental stages. Solid gray lines represent fold change values calculated from RNA-seq data, while dashed red lines represent fold change values measured by qRT-PCR. Error bars indicate the standard deviation (n = 3) of the qRT-PCR results.

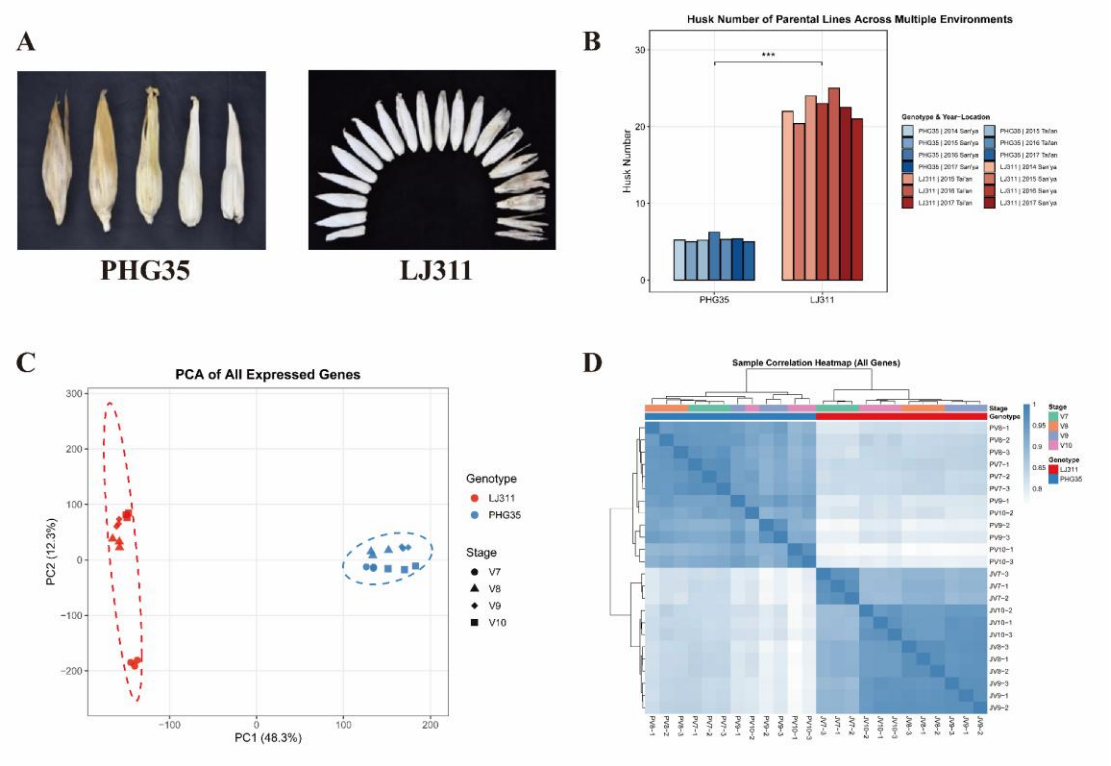


Figure 1. Phenotypic difference and global transcriptomic characterization of husk number. (A) Mature husk phenotypes of the two parental lines. Left, PHG35; right, LJ311. (B) Mean husk number of PHG35 and LJ311 across multiple years and locations. (C) PCA of the 24 samples based on all genes passing the basic expression threshold. Colors indicate genotypes (LJ311, red; PHG35, blue), and shapes indicate developmental stages (V7, circle; V8, triangle; V9, diamond; V10, square). (D) Pearson correlation heatmap with hierarchical clustering of the 24 samples. Darker blue indicates higher correlation.

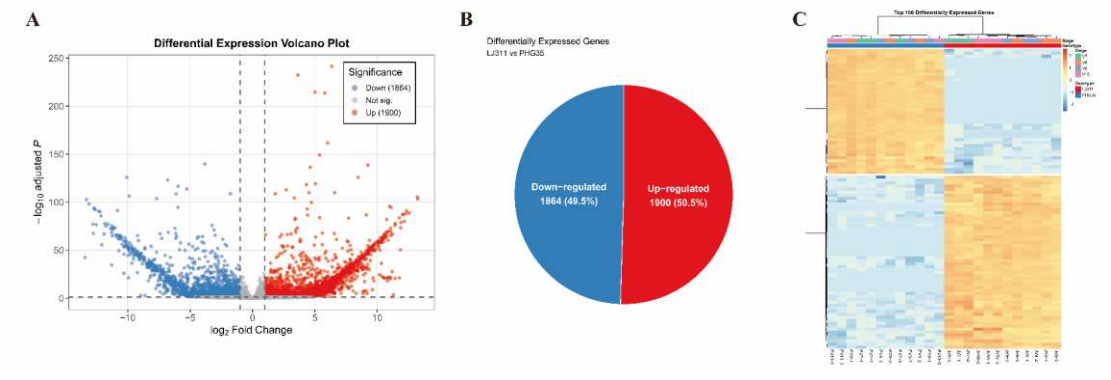


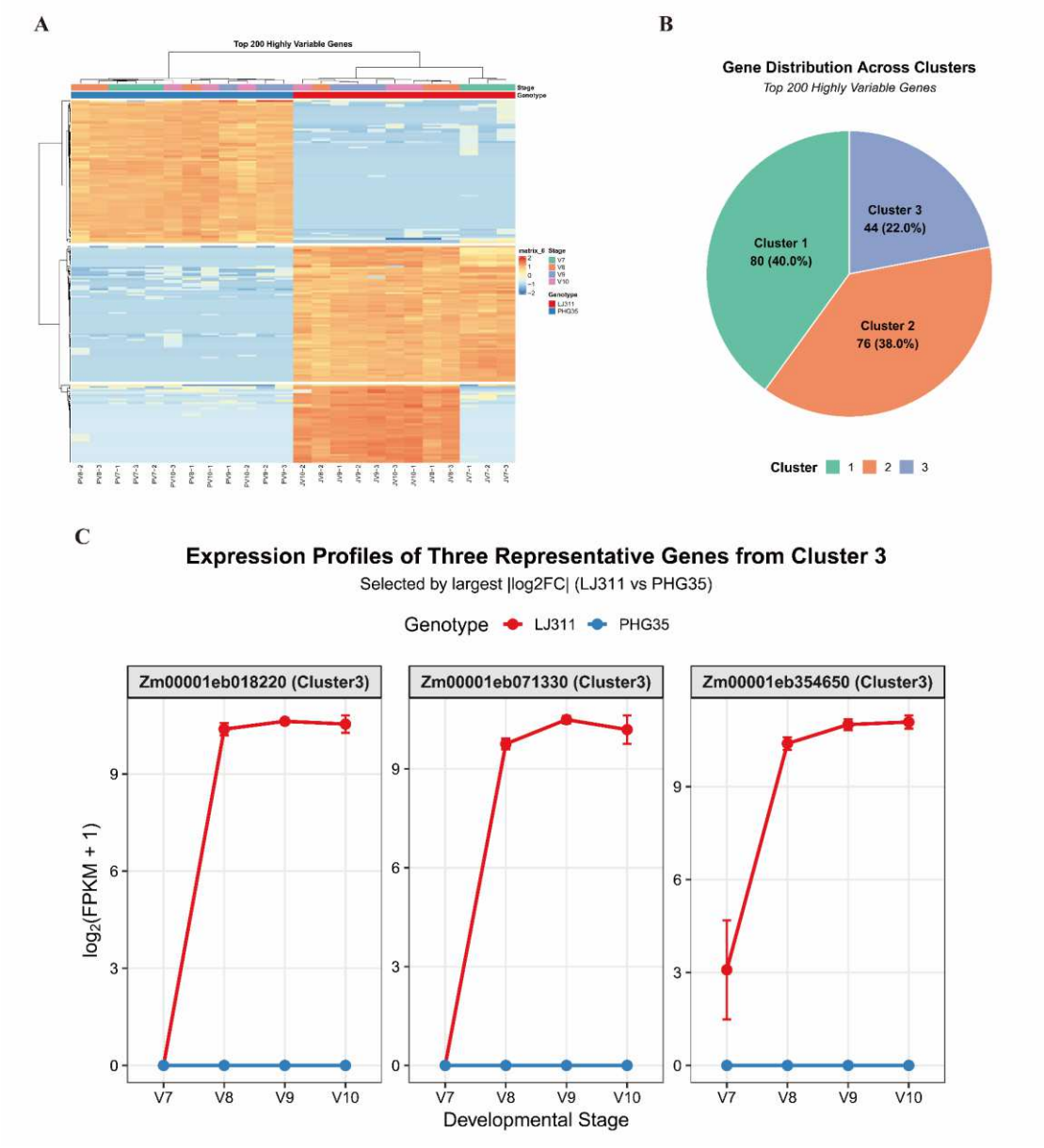
Figure 2. Identification and expression profiling of differentially expressed genes (DEGs)

690 **between LJ311 and PHG35.** (A) Volcano plot of DEGs. The screening thresholds were set at

691 |log2FC| > 1 and padj < 0.05. Red, up-regulated; blue, down-regulated; gray, not significant. (B) Pie

692 chart showing the proportion of up-regulated and down-regulated DEGs. (C) Heatmap of the top

693 100 DEGs. Red indicates high expression, and blue indicates low expression.



694

695 **Figure 3. Unsupervised hierarchical clustering analysis of highly variable genes.** (A) Heatmap

696 of the top 200 most variable genes across all 24 samples. Genes were Z-score normalized and

697 clustered into three major groups (Cluster 1–3) based on their expression patterns. (B) Pie chart

showing the number of genes in each of the three clusters. (C) Expression dynamics of three representative genes from Cluster 3 across developmental stages and genotypes. These three genes had the largest |log2FC| values within the cluster. Data are presented as mean ± SE.

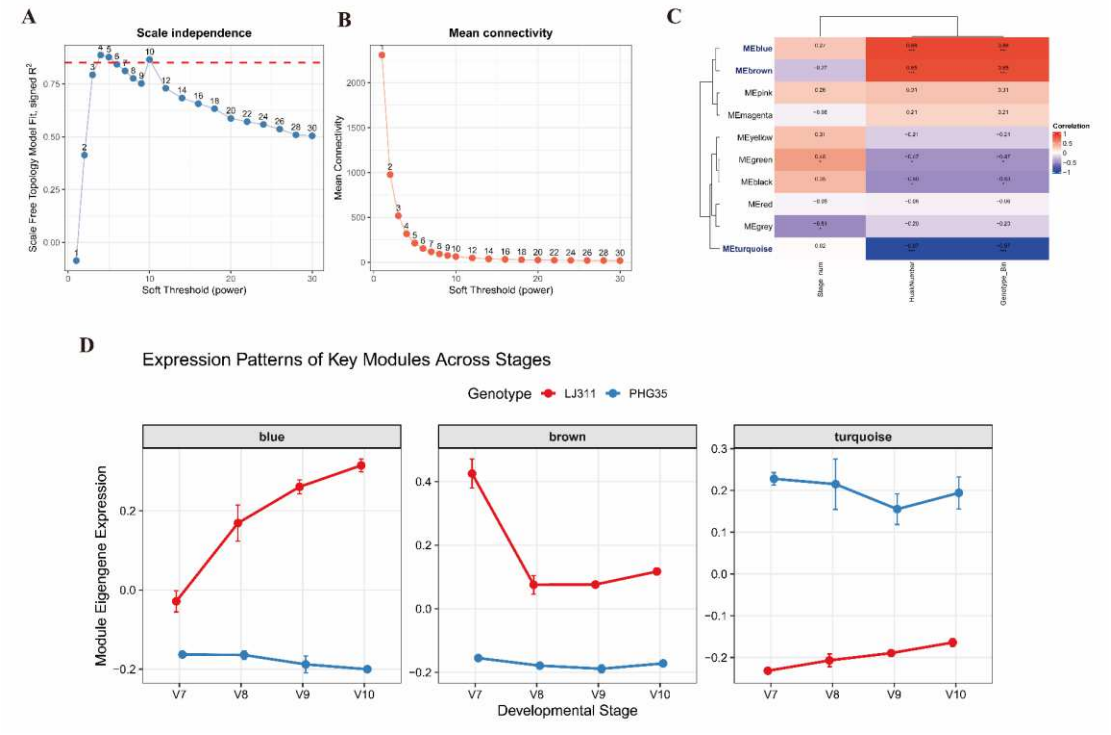


Figure 4. WGCNA of husk number-related modules. (A, B) Scale independence (A) and mean connectivity (B) across soft-thresholding powers. The red line in (A) marks $R^2 = 0.85$. (C) Module–trait correlation heatmap. Blue, brown, and turquoise modules were most significantly associated with husk number. *** $p < 0.001$. (D) Expression trends of module eigengenes for the three modules across developmental stages. Red, LJ311; blue, PHG35. Data are mean ± SE.

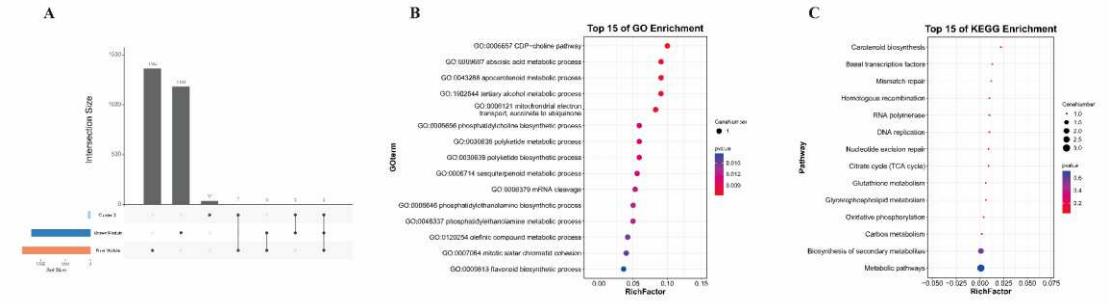


Figure 5. Cross-validation between hierarchical clustering and WGCNA. (A) UpSet plot

showing the overlap of genes between Cluster 3 and the blue and brown modules. (B) GO enrichment bubble plot of all 44 genes in Cluster 3. Bubble size represents the number of enriched genes, and bubble color represents the adjusted *p*-value (p_{adj}). (C) KEGG enrichment bubble plot of all 44 genes in Cluster 3. Bubble size represents the number of enriched genes, and bubble color represents the adjusted *p*-value (p_{adj}).

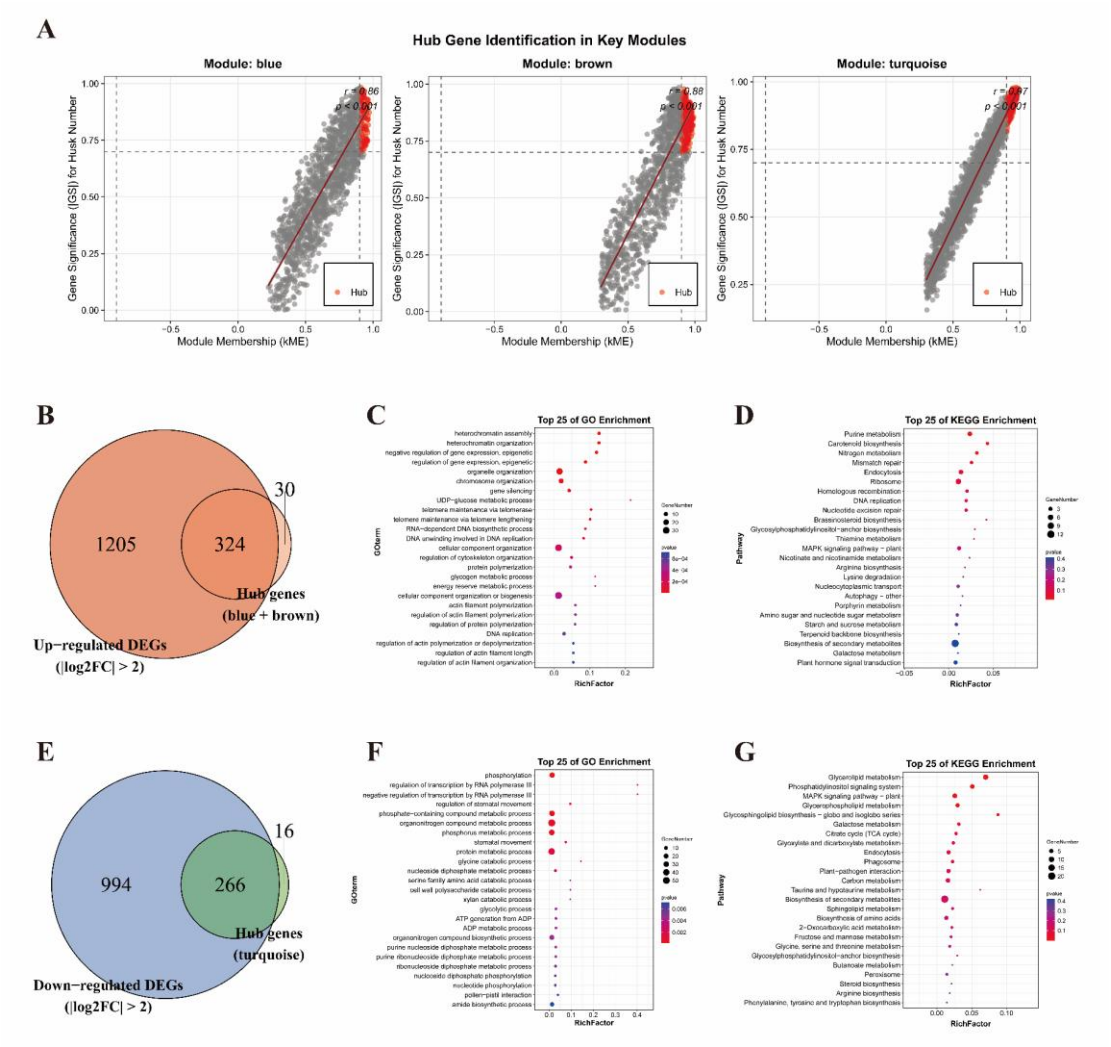


Figure 6. Hub gene identification and functional enrichment of high-confidence candidates. (A) kME–GS scatter plots for the blue, brown, and turquoise modules. Dashed lines mark |kME| = 0.9 and |GS| = 0.7; red dots, hub genes. (B, E) Venn diagrams showing intersections between hub genes and DEGs (|log2FC| > 2) for positively (B) and negatively (E) correlated modules. (C, D) GO

(C) and KEGG (D) enrichment bubble plots of up-regulated candidates. (F, G) GO (F) and KEGG (G) enrichment bubble plots of down-regulated candidates. For all bubble plots, bubble size indicates gene count, and color indicates adjusted *p*-value (p_{adj}).

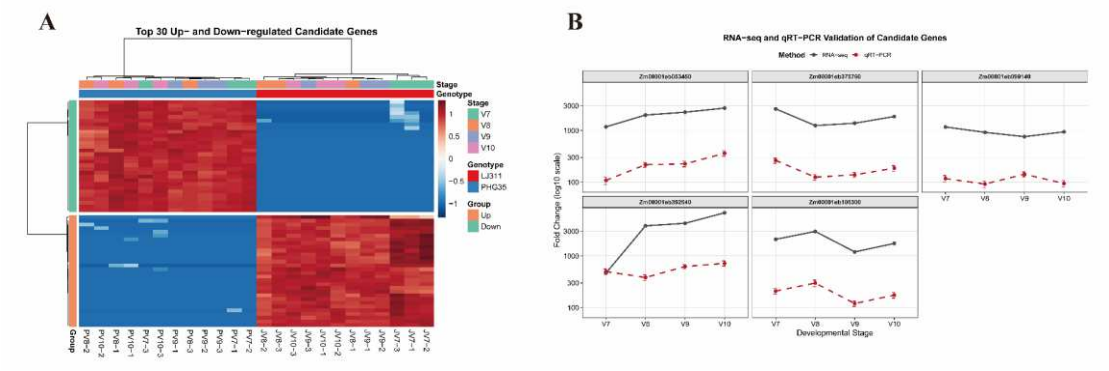


Figure 7. Expression patterns and experimental validation of final candidate genes. (A) Heatmap of the top 30 up- and down-regulated candidate genes. Genes are clustered into two groups based on their regulatory directions. (B) Fold change comparison between RNA-seq and qRT-PCR for five representative hub genes across four developmental stages. Solid gray lines represent fold change values calculated from RNA-seq data, while dashed red lines represent fold change values measured by qRT-PCR. Error bars indicate the standard deviation (n = 3) of the qRT-PCR results.

Supplementary Files

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

- TableS1.xlsx
- TableS5.xlsx
- FigureS1.tif
- TableS2.xlsx
- FigureS4.tif
- TableS3.xlsx
- FigureS2.tif
- FigureS3.tif
- FigureS5.tif
- TableS4.xlsx