

Research Article

Open Access



CrossMark

<sup>1</sup> Department of Genetic Resources, Desert Research Center (DRC), Cairo, Egypt.

\* To whom correspondence should be addressed: [drmokhtarsaid@yahoo.com](mailto:drmokhtarsaid@yahoo.com)

**Editor:** Aladdin Hamwih, *International Center for Agricultural Research in the Dry Areas (ICARDA), Giza, Egypt.*

**Reviewer(s):**

Clara R. Azzam, *Cell Research Department, Field Crops Research Institute, Agricultural Research Center, Giza 12619, Egypt.*

Fida Alo, *International Center for Agricultural Research in the Dry Areas (ICARDA), Beirut, Lebanon.*

Thoria Rashad Mohamed, *Cytology and genetics, Botany department, women faculty, Cairo, Egypt.*

**Received:** July 10, 2023

**Accepted:** Sep 6, 2023

**Published:** September 15, 2023

**Citation:** Rizk MS. Unveiling the architecture variability of coding regions in *Prunus persica*: Identification and characterization of key gene families impacting plant development. 2023 Sep 15;6:bs202302


**Copyright:** © 2023 Rizk. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and supplementary materials.

**Funding:** The authors have no support or funding to report.

**Competing interests:** The author declare that they have no competing interests.

## Unveiling the architecture variability of coding regions in *Prunus persica*: Identification and characterization of key gene families impacting plant development

Mokhtar Said Rizk <sup>1</sup> 

### Abstract

This study aimed to investigate gene families related to *Prunus persica* (peach) and their evolutionary relationships. Gene identification and sequence alignment methods were used to characterize Homeobox, MADS-box, NAC, and Receptor-kinase-like gene families. Gene expression data analysis was conducted to assess gene activity and regulatory mechanisms in peach plant development. Gene identification involved a local BLASTp analysis using protein sequences from *Arabidopsis thaliana* gene families as queries against the peach proteome. Multiple sequence alignment and phylogenetic tree construction were performed using ClustalW and iTOL, respectively. Gene expression analysis utilized the GEO2R online tool and focused on two datasets, GSE71561 and GSE71460. Gene identification revealed 593 genes in total, with varying gene counts and sequence lengths among the different families. Phylogenetic analysis demonstrated distinct clusters within each gene family. Protein-protein interaction analysis showed significant interconnectivity among Homeobox transcription factors, while MADS-box and NAC gene families displayed lower connectivity. Enrichment analysis highlighted developmental processes associated with these gene families. Gene expression analysis provided insights into peach fruit development, ripening, and postharvest changes, revealing differential gene expression patterns and regulatory mechanisms. This study provides valuable insights into the gene families related to *Prunus persica*, their evolutionary relationships, and regulatory mechanisms. The identified genes and their functional characteristics contribute to a better understanding of peach development and fruit phenotypes. Further research is required to elucidate the implications of gene variations and interactions in the studied system.

**Keywords:** Peach, Gene expression analysis, Gene families, MADS-box, NAC gene family, Receptor-kinase-like

### Introduction

Peaches, scientifically known as *Prunus persica*, are fruit trees belonging to the rose family (*Rosaceae*) and are cultivated in warmer temperate regions of both the Northern and Southern hemispheres [1]. They are widely enjoyed fresh and are commonly used in baking, such as in pies and cobblers. Canned peaches are also a popular staple in many areas. Varieties with yellow flesh are particularly notable for their high vitamin A content [2]. Compared to some other fruit trees, peach trees have a relatively shorter lifespan. In certain regions, orchards are replanted every 8 to 10 years, while in others, trees may continue to produce satisfactorily for 20 to 25 years or even longer, depending on their resistance to diseases, pests, and winter-related damage [3].

Countries such as Spain, Italy, the United States, Chile, and Australia consider peach as highly significant *Prunus* specie. The European Union holds the second position worldwide in peach production, with China being the only country ahead. Between 2018 and 2020, the EU averaged an annual production of 3612,000 tons, while the harvested area for peaches in 2019 encompassed 206,660 hectares. Leading the rankings, Spain boasts 77,464 hectares of peach cultivation and produces 1480,000 tonnes per year, closely followed by Italy and Greece [4].

Homeobox genes, MADS-box genes, NAC genes, and receptor-kinase-like genes are all involved in the regulation of gene expression and play important roles in the development and growth of organisms. Homeobox genes, MADS-box genes, NAC genes, and receptor-kinase-like genes are crucial players in gene expression regulation and exert significant roles in organismal development and growth [5; 6]. Specifically, homeobox genes involved in floral organ development have been found to interact with MADS-box genes, which in turn govern the expression of a wide array of genes responsible for flower development and organ identity, including other MADS-box genes and homeobox genes [7]. On the other hand, NAC genes exhibit interactions with both homeobox genes and MADS-box genes, thereby influencing their expression and contributing to the regulation of plant development and stress responses [8; 9].

Receptor-kinase-like genes represent a diverse gene group encoding receptor-like kinases (RLKs) or receptor-like proteins, and some RLKs have demonstrated interactions with transcription factors, including members of the MADS-box and NAC gene families. These interactions are involved in the regulation of gene expression and signaling pathways crucial for various developmental processes and stress responses [10; 11]. Collectively, the intricate network of interactions and regulatory mechanisms within these gene families contributes to the meticulous control of gene expression and the coordination of developmental processes in organisms [12].

Studying these gene families, including Homeobox genes, MADS-box genes, NAC genes, and receptor-kinase-like genes, is of great importance in the field of peach genomics. These gene families are involved in the regulation of gene expression and play significant roles in the development and growth of organisms [13]. By understanding the functions and interactions of these genes, researchers can gain insights into the molecular mechanisms underlying peach development and improve their understanding of the genetic factors that influence important traits in peach [14].

The aim of this study is to accomplish the following objectives. Firstly, our aim is to identify and annotate the gene family members, which include Homeobox genes, MADS-box genes, NAC genes, and receptor-kinase-like genes, within the peach genome. This objective will involve conducting a comprehensive genome-wide analysis and utilizing bioinformatics tools to accurately identify and classify these gene family members. Secondly, we aim to characterize the gene family members to gain insights into their diversity and functional variations. By investigating their unique features and roles in peach, we can enhance our understanding of the genetic makeup of these genes. Lastly, we intend to investigate the expression patterns of the gene family members throughout plant development. This will be achieved through high-throughput transcriptomic analyses, such as RNA-sequencing, to capture the dynamic expression profiles of these genes. By accomplishing these objectives, our study aims to contribute to a deeper understanding of the genetic

characteristics, functional diversity, and expression dynamics of Homeobox genes, MADS-box genes, NAC genes, and receptor-kinase-like genes in peach.

## Methods

### Gene identification and sequence alignment

The methodology employed in this study aimed to investigate the gene families related to *Prunus persica* (peach) and their evolutionary relationships. The genome data used for the analysis was obtained from *Prunus persica* NCBIv2 [15], as provided by the NCBI RefSeq database.

The gene families of interest included Homeobox, MADS-box, NAC, and Receptor-kinase-like. The gene set comprised a total of 593 genes, while Homeobox, MADS-box, NAC, and Receptor-kinase-like gene families consisted of 91, 106, 94, and 302 genes, respectively. To further analyze these gene families, information from *Arabidopsis thaliana* gene families was downloaded from the TIAR database, providing valuable reference data [16].

To identify the corresponding gene family members in the peach genome, a local BLASTp analysis was conducted. The protein sequences from the *Arabidopsis thaliana* gene families were used as queries against the peach proteome. The BLASTp analysis was performed using the following parameters: "-outfmt 6 -evalue 1e-5 -num\_threads 4" [17]. Genes that showed less than 200 base pairs of match were subsequently filtered out from further analysis.

To assess the evolutionary relationships among the identified gene family members, a multiple sequence alignment was performed using ClustalW, a widely used tool for aligning multiple protein or nucleotide sequences [18]. The resulting alignment was then utilized to construct a phylogenetic tree. The phylogenetic tree construction was accomplished using iTOL (Interactive Tree Of Life), an interactive web-based tool specifically designed for visualizing and manipulating phylogenetic trees [19]. In the analysis of sequencing data, the Alignstatplot tool [20], was employed to showcase both shared and distinct genomic regions.

By implementing these methods, the study aimed to explore the gene families related to *Prunus persica* and gain insights into their evolutionary relationships. The genome data from *Prunus persica* NCBIv2, combined with gene families information from *Arabidopsis thaliana*, enabled the identification and characterization of the gene family members in peach. The BLASTp analysis, multiple sequence alignment, and phylogenetic tree construction provided valuable information regarding the evolutionary history and relationships of these gene families.

### Gene expression data analysis

Gene expression analysis was conducted using the GEO2R online tool [21] to assess gene activity and regulatory mechanisms in peach plant development. The researchers focused on gene families such as Homeobox Transcription Factor, MADS-

box Gene, NAC Transcription Factor, and Receptor kinase-like Gene families.

Two datasets, GSE71561 [22] and GSE71460 [23], were analyzed. GSE71561 aimed to understand the molecular control of peach fruit development, including interactions between seed and mesocarp. It comprised 39 samples representing various developmental stages, covering flower, seed, and mesocarp tissues. Each sample had three biological replicates. Gene enrichment analysis identified differentially expressed genes within the same tissue stage, and the researchers examined tissue specificity versus stage specificity of the regulation.

GSE71460 focused on investigating the effects of gibberellic acid application and cold storage on postharvest textural changes in peaches, specifically woolliness. The dataset included 12 samples with four treatments, each having three biological replicates. Approximately half of the investigated genes were analyzed, revealing significant differential expression. Gene ontology and set enrichment analyses indicated the involvement of cellular processes, developmental processes, and complex transcriptional responses related to cell wall metabolism, hormone metabolism, and signaling in response to cold storage and gibberellic acid.

By using the GEO2R online tool and analyzing these datasets, the researchers gained insights into gene expression patterns and regulatory mechanisms underlying peach fruit development, ripening, and postharvest changes. This study contributes to a better understanding of the molecular processes governing these complex phenotypes in peaches. Protein-protein interaction (PPI) analysis was performed using the STRING database [24] and enrichment analysis. The PPI network was visualized using the Cytoscape program [25].

## Results and Discussion

### Gene identification and sequence variationa

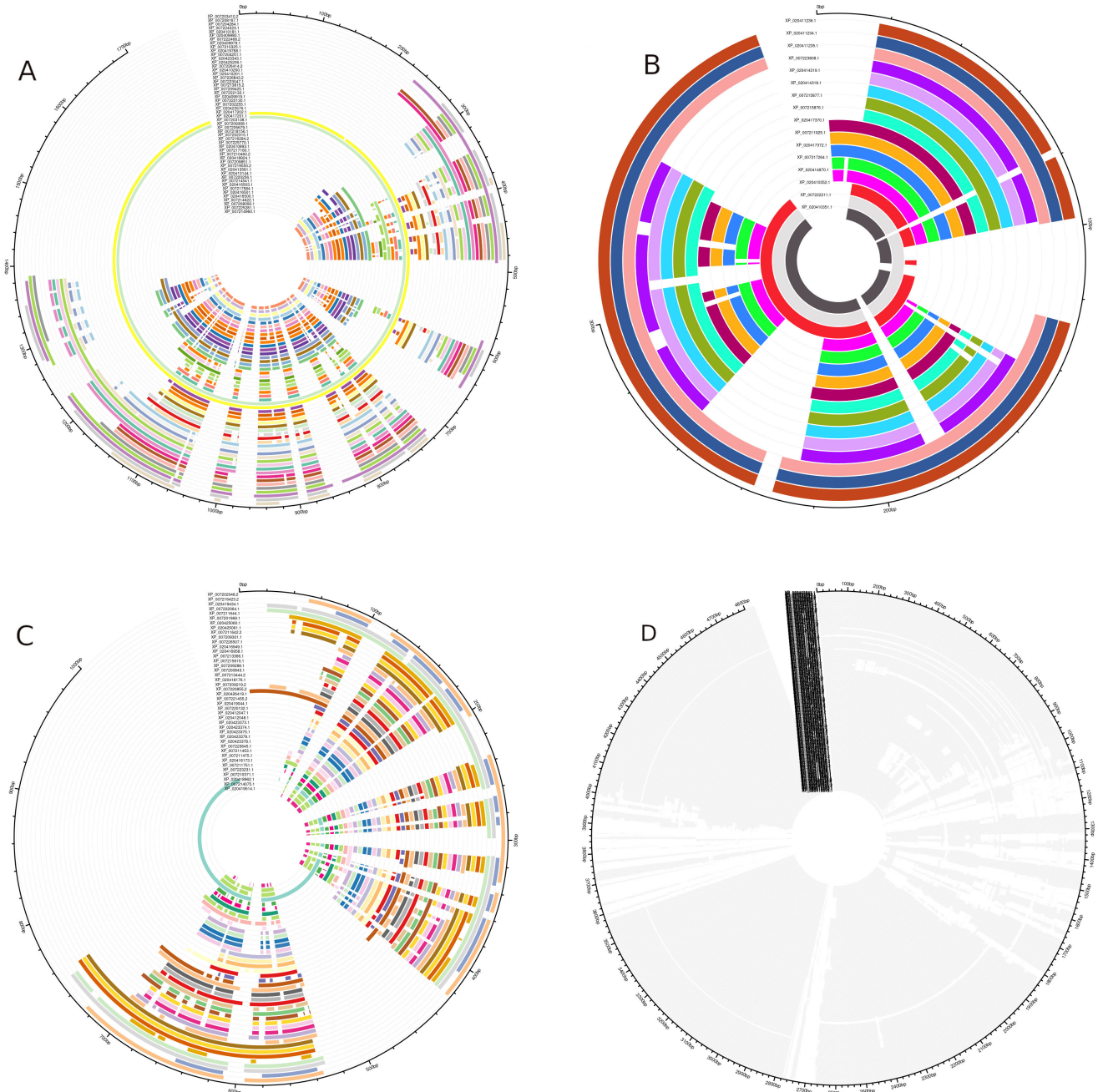
Genes belonging to four distinct gene families were identified across peach genome (**Figure 1**). The results of the analysis provided valuable insights into the gene counts and characteristics of each family. The Homeobox transcription factor family consisted of 52 genes, with a total length of 33,042 base pairs (bp). Among these genes, the maximum sequence length observed was 837 bp, while the minimum sequence length was 1,764 bp. For the MADS-box Gene family, a total of 16 genes were identified, spanning a length of 4,522 bp. The maximum sequence length observed in this family was 240 bp, and the minimum sequence length was 349 bp. In the NAC transcription factor family, a total of 42 genes were identified, with a combined length of 15,903 bp. The maximum sequence length observed within this family was 591 bp, while the minimum sequence length was 862 bp (**Figure 1**). Finally, the Receptor kinase-like Gene family exhibited the largest gene count, with 366 genes identified. These genes spanned a length of 300,904 bp. The maximum sequence length observed in this family was 626 bp, and the minimum sequence length was 556 bp (**Figure**

1). These results highlight the variations in gene counts and sequence lengths among the different gene families. They provide valuable information about the diversity and characteristics of genes within each family, shedding light on their potential roles and functions in the studied system.

The multiple sequence alignment conducted in this study provided valuable insights into the patterns of sequence variation within the gene families analyzed (**Figure 1**). These variations were evident through the presence of sequence gaps observed in the alignments. It was noticeable that certain sequences exhibited substantial portions of amino acid sequences that were absent in others, indicating potential differences in gene function and underlying mechanisms (**Figure 1**). The presence of sequence gaps in the multiple sequence alignments suggests the occurrence of insertions or deletions in the gene sequences among different members of the gene families. Such variations in sequence length and composition could have significant implications for the functionality and regulation of these genes. The observed differences in amino acid sequences may lead to variations in protein structure and function, potentially influencing their roles within the studied system [26]. By identifying these sequence variations, this study contributes to a deeper understanding of the diversity and potential functional divergence within the gene families analyzed. Further investigations into the specific functions and mechanisms associated with the unique sequence features observed could provide valuable insights into the evolutionary dynamics and functional specialization of these genes [27]. Additionally, the phylogenetic analysis based on the sequence alignments revealed distinct clusters of genes within each family, suggesting differences in gene functions and evolutionary relationships (**Figure 2**).

The comparison between the results obtained in this study and previous studies provides valuable insights into the gene counts and characteristics of different gene families. A total of 73 homeobox-like genes in the grapevine genome were identified in a previous study [28]. The comparison between the results obtained in this study and previous studies provides valuable insights into the gene counts and characteristics of different gene families. In the current study, our analysis identified 52 genes associated with Homeobox Transcription Factor Family, suggesting variations in the composition and abundance of Homeobox Transcription Factor genes between species or datasets. Similarly, in the MADS-box Gene Family, our analysis identified only 16 genes, indicating potential disparities in the MADS-box gene repertoire between the studied system and previous studies in rice [29]. In relation to the NAC Transcription Factor Family, our analysis identified 42 genes, contrasting with the 140 putative NAC or NAC-like genes identified in rice [30]. These contrasting gene counts suggest variations in the presence or abundance of NAC genes between the studied system and rice. The specific gene counts mentioned in the provided articles were higher for the Homeobox Transcription Factor Family (73 genes) and MADS-box Gene Family (75 genes) in grapes and rice, re-



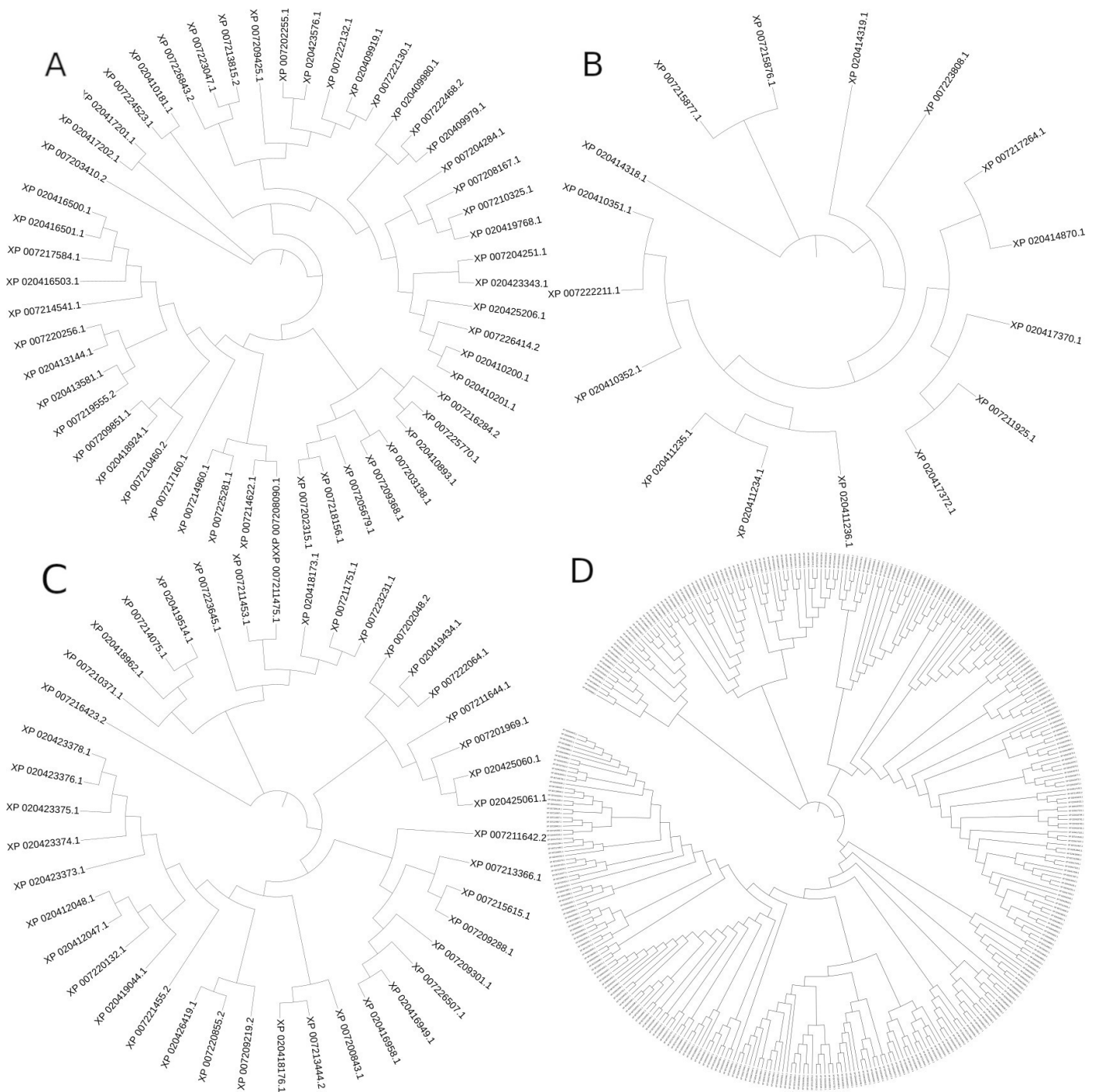


**Figure 1.** Multiple sequence alignment of identified sequences from the Homeobox transcription factor family (A), MADS-box Gene family (B), NAC transcription factor family (C), and Receptor kinase-like Gene family (D) in the peach genome. Sequences are color-coded, and gaps represent missing regions in the alignment.

spectively. However, the gene counts obtained in this study may be specific to the analysis or dataset used, and further research is required to elucidate the reasons behind these variations and their implications in the studied system.

**Protein-protein interaction analysis**

The analysis of the Homeobox transcription factor family in the peach genome using protein-protein interaction (PPI) network analysis revealed significant interconnectivity and potential functional relationships among the proteins (**Figure 3**). The



**Figure 2.** Phylogenetic trees illustrating the evolutionary relationships of identified sequences from the Homeobox transcription factor family (A), MADS-box Gene family (B), NAC transcription factor family (C), and Receptor kinase-like Gene family (D) in the peach genome. The trees provide insights into the genetic divergence and relatedness among the sequences within each gene family, highlighting their evolutionary history and potential functional similarities or distinctions.

network consisted of 25 proteins and 61 interactions, indicating a complex regulatory network involved in peach development. The average local clustering coefficient suggested clustering of nodes, further supporting the notion of functional co-

operativity among the Homeobox transcription factors. The enriched Gene Ontology (GO) terms associated with the Homeobox transcription factor family included various processes related to key developmental processes and organ polarity estab-

ishment in peach, such as determination of bilateral symmetry, primary shoot apical meristem specification, xylem development, meristem initiation, and polarity specification of the adaxial/abaxial axis. These findings highlight the important roles of Homeobox transcription factors in orchestrating developmental processes and regulating gene expression and cellular processes in peach.

Several overlapping themes can be observed when comparing these findings with published studies. A previous study focused on HD-Zip III homeobox genes and their involvement in vascular differentiation [31]. They identified a new member of the HD-Zip III genes, ZeHB-13, which exhibited restricted expression in the procambium. This finding aligns with the enriched GO term of xylem development identified in the peach Homeobox transcription factor family. Both studies suggest that Homeobox transcription factors play crucial roles in the regulation of vascular development. The investigation of the TDIF peptide signaling pathway and its impact on vascular stem cell proliferation in Arabidopsis revealed the WOX4 homeobox gene as a key target of the TDIF signaling pathway, promoting the proliferation of procambial/cambial stem cells [32]. This finding is consistent with the enriched GO term of meristem initiation and the regulation of primary shoot apical meristem specification observed in the peach Homeobox transcription factor family. Both studies imply the involvement of Homeobox transcription factors in the regulation of stem cell proliferation and meristem development. Furthermore, a study on homeobox genes associated with lignification in bamboo shoots identified 115 homeobox genes, including several classes such as the KNOX class, that interacted with other transcription factors involved in lignin synthesis [33]. This interaction network parallels the potential functional relationships observed in the protein-protein interaction network analysis of the peach Homeobox transcription factor family. Additionally, the upregulation of homeobox genes in shoots as the height increased suggests their involvement in developmental processes similar to the enriched GO terms identified in the peach study.

Similarly, the MADS-box gene family in the peach genome was analyzed using PPI network analysis (Figure 3). The network showed relatively low interconnectivity with 6 proteins and 5 interactions. The average local clustering coefficient suggested some tendency for nodes to cluster together. However, the PPI enrichment analysis revealed functional associations or co-regulation among the MADS-box gene family members. The functional enrichment analysis identified processes such as Pollen maturation, Regulation of pollen tube growth, Plant ovule development, and Flower development associated with the MADS-box gene family. These findings indicate their involvement in important processes like pollen development, fertilization, and flower formation in peach. In the context of these results, an investigation into the functional conservation of MADS-box genes in Arabidopsis and rice pollen maturation was conducted [34]. The findings revealed that the disruption of specific MIKC-type

genes in rice resulted in severe defects in pollen maturation and germination, providing supporting evidence for the involvement of MADS-box genes in pollen development. Furthermore, a study on MADS-box gene evolution beyond flowers identified novel MADS-box genes in Arabidopsis, indicating that the evolution of the MADS-box family involved rapid and simultaneous functional diversification in both vegetative and reproductive structures [35]. These findings align with the present research, demonstrating functional associations or co-regulation among the MADS-box gene family members. Additionally, a study focused on the pineapple MADS-box gene family and its evolutionary history emphasized the importance of understanding the ancestral form of monocot flowers and their development, further supporting the relevance of studying MADS-box genes in different plant species [36].

In this study, we analyzed the NAC transcription factor family in the peach genome using PPI network analysis and enrichment analysis (Figure 3). The PPI network analysis revealed a low level of connectivity and clustering within the NAC transcription factor network, suggesting potential complexity in their regulatory mechanisms. Additionally, the lack of significant enrichment in the PPI analysis indicated sparse or transient interactions among these proteins. However, the functional enrichment analysis provided valuable insights into the involvement of NAC transcription factors in diverse plant development processes.

Our findings are consistent with previous studies that have investigated the role of NAC transcription factors in specific plant developmental processes. For instance, a study demonstrated that the VASCULAR-RELATED NAC-DOMAIN7 (VND7) transcription factor regulates the expression of genes required for xylem vessel element formation in Arabidopsis. Multiple transcription factors, including VND1-VND7, were identified as putative positive regulators of VND7 expression. This study highlights the importance of NAC transcription factors in xylem vessel differentiation, which aligns with our functional enrichment analysis showing their involvement in xylem vessel differentiation and secondary cell wall biogenesis [37].

Another study discovered that VND-INTERACTING2 (VNI2), a NAC domain transcription factor, negatively regulates xylem vessel formation in Arabidopsis. VNI2 interacts with VND7 and other VND family proteins, acting as a transcriptional repressor and repressing the expression of vessel-specific genes regulated by VND7. This finding supports our findings of the potential complexity in the regulatory mechanisms of NAC transcription factors and their role in xylem vessel differentiation [38]. Furthermore, a study investigated the function of XYLEM NAC DOMAIN1 (XND1) in xylem differentiation. It revealed that XND1 interacts with the RETINOBLASTOMA-RELATED (RBR) protein and acts as a transcriptional repressor to inhibit differentiation. These findings provide insights into the inhibitory role of NAC transcription factors in xylem differentiation, which aligns with our functional enrichment analysis showing their involvement in secondary cell wall biogenesis [39]. In addi-



tion to their role in vascular development, NAC transcription factors have been implicated in other plant developmental processes. For example, a study investigated the involvement of NAM genes, regulated by miR164, in floral-boundary morphogenesis in tomato. The NAM gene GOBLET was identified as a key regulator of floral-boundary formation. Our functional enrichment analysis revealed the involvement of NAC transcription factors in shoot meristem specification and root cap development, highlighting their role in various aspects of plant growth and development [40]. Another study investigated the role of CUP-SHAPED COTYLEDON1 (CUC1), a NAC transcription factor, in the establishment of shoot organ boundaries in Arabidopsis. It was found that CUC1 activates the expression of LIGHT-DEPENDENT SHORT HYPOCOTYLS 4 (LSH4) and LSH3, members of the ALOG gene family, in shoot organ boundary cells. This study provides insights into the regulatory mechanism of shoot organ boundary establishment mediated by NAC transcription factors [41].

The Receptor kinase-like gene family in the peach genome was analyzed using PPI network analysis and enrichment analysis (Figure 3). The PPI network consisted of 151 nodes and 966 edges, indicating a dense network of protein interactions. The network exhibited a high level of interconnectivity, suggesting functional relevance and potential cooperativity among the genes. The functional enrichment analysis provided insights into the involvement of the Receptor kinase-like gene family in various developmental processes critical for plant growth and response to environmental stimuli. The analysis also highlighted molecular functions related to signal transduction and receptor binding, emphasizing the importance of receptor-mediated signaling pathways in peach. References to relevant scientific publications covered topics such as genetic incompatibility [42], self-incompatibility signaling [43], protein secretion, gene duplications, and orthologous relationships [44], providing further insights into the functional characteristics and evolutionary aspects of the Receptor kinase-like gene family in peach. The enrichment analysis included classifications from various databases and ontologies, offering additional functional annotations and domain information associated with the Receptor kinase-like gene family, enhancing our understanding of their functional properties and characteristics.

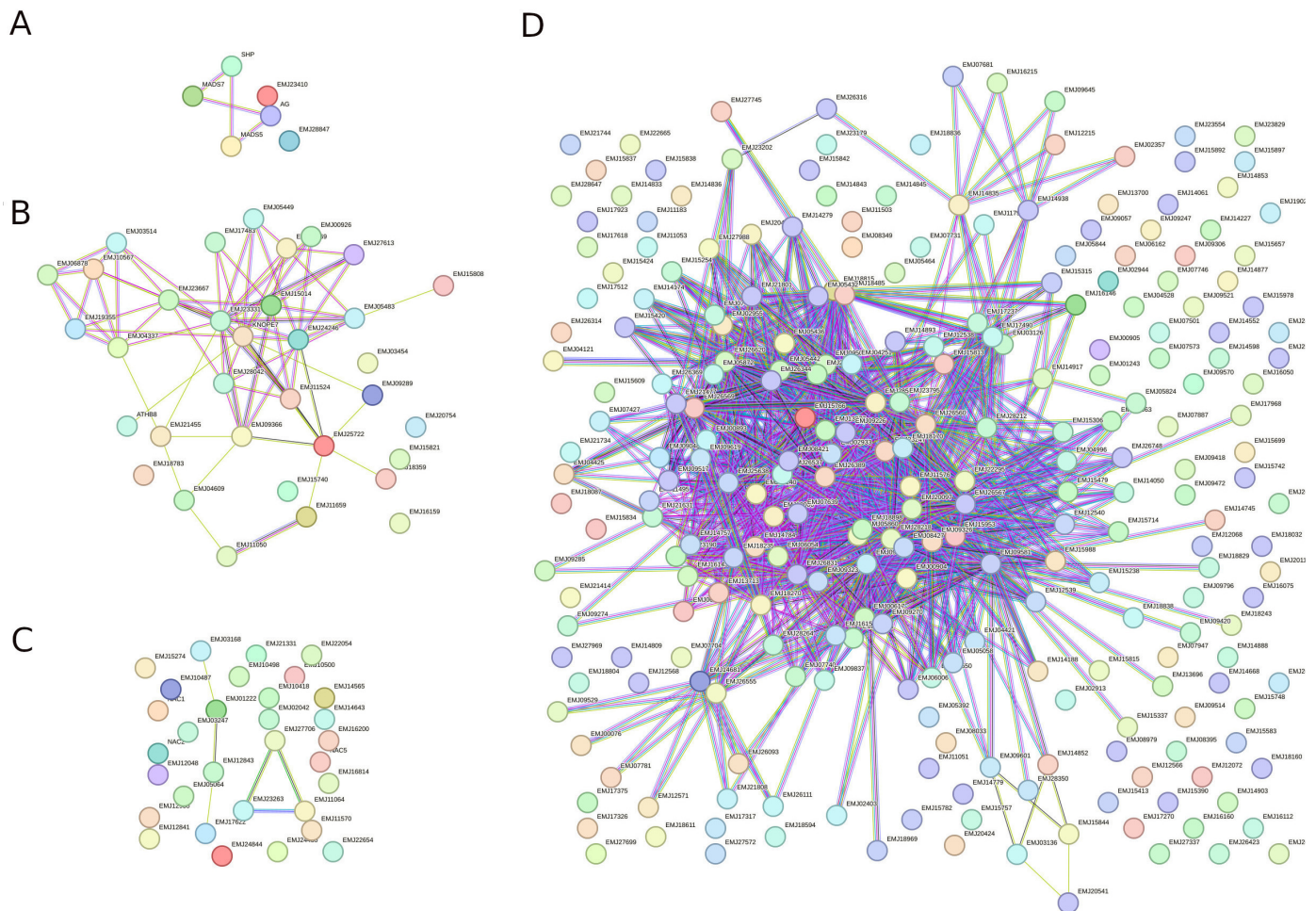
### Gene differential expression analysis

Gene expression analysis was performed to identify genes highly associated with peach development using two different datasets: GSE71460 and GSE71561. In GSE71460, which focused on the effects of gibberellic acid application and cold storage on postharvest textural changes in peaches, several important genes were identified. These included SKP1-like, F-box/FBD/LRR-repeat protein, leucine-rich repeat receptor-like, Triose phosphate isomerase cytosolic isoform-like protein, and UDP-glycosyltransferase 87A1-like. On the other hand, GSE71561 aimed to elucidate the molecular regulation of peach fruit de-

velopment, particularly the interactions between the seed and mesocarp tissues. In this dataset, isoleucine N-monooxygenase 2, myb-related protein, and transcription factor bHLH75-like were among the most important genes identified. Enrichment analysis was conducted on both datasets to gain insights into the functional significance of the identified genes. In GSE71460, the enrichment analysis revealed several enriched terms and processes that suggested the involvement of these genes in various cellular processes, metabolic activities, and molecular interactions related to peach development.

In a previous study, the expression patterns of Arabidopsis-SKP1-like (ASK) genes, which play important roles in ubiquitin-mediated proteolysis and various biological processes, were examined [45]. The findings of that study shed light on the expression profiles of SKP1-like genes in plants. Our study, focusing on gene expression during fruit development, shares similarities with the previous study. Therefore, our results contribute to understanding the roles of SKP1-like genes in peach development. Another investigation explored the molecular mechanisms of self-incompatibility in plants. It identified a Skp1-like protein as a crucial component of a protein complex involved in pollen compatibility [46]. Although the study's focus was on self-incompatibility, the importance of Skp1-like proteins in plant reproductive processes can provide insights into peach fruit development. Additionally, we observed differential expression of triose phosphate isomerase, an essential enzyme in the glyoxalase pathway, during peach development. This finding aligns with a separate study that discussed the implications of triose phosphate isomerase and the glyoxalase pathway in plant responses to abiotic stress and signaling [47]. While the latter study primarily examined stress responses, their insights into triose phosphate isomerase activity are valuable for understanding its involvement in peach development. By comparing our findings with the mentioned studies, we enhance the current understanding of gene expression patterns during peach development and establish potential connections to processes such as ubiquitin-mediated proteolysis, self-incompatibility, and stress responses.

In the case of GSE71561, the enrichment analysis highlighted several important gene-enriched pathways. These included cellular anatomical entity, intracellular, catalytic activity, membrane-bounded organelle, intracellular membrane-bounded organelle, metabolic process, cellular process, response to stimulus, organelle, intracellular organelle, chloroplast, cytoplasm, plastid, binding, thylakoid, transit peptide, chloroplast thylakoid, response to stress, cellular metabolic process, membrane, developmental process, biosynthesis of secondary metabolites, oxidation-reduction process, thylakoid, and stromule, metabolic pathways, photosystem, thylakoid, plastid envelope, chloroplast envelope, oxidoreductase activity, response to abiotic stimulus, anatomical structure development, thylakoid membrane, and chloroplast thylakoid membrane. These enriched pathways provide insights into the molecular processes and organelle-specific functions as-



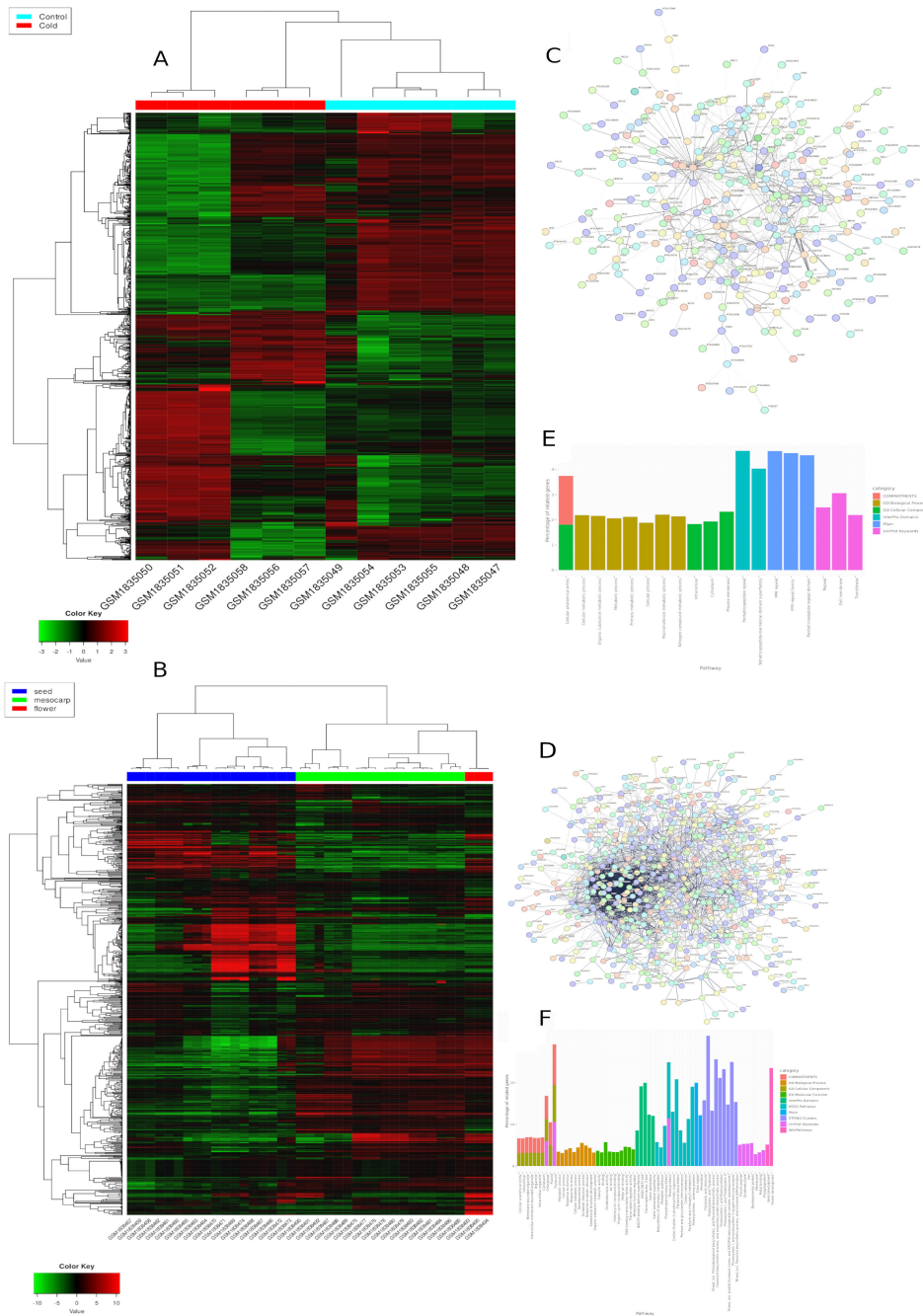
**Figure 3.** Protein-protein interaction networks of the identified sequences from the MADS-box gene family (A), Homeobox transcription factor family (B), NAC transcription factor family (C), and Receptor kinase-like gene family (D) visualized using the STRING database. The networks provide a comprehensive overview of the protein interactions within each gene family, shedding light on potential functional associations and regulatory relationships among the identified proteins.

sociated with peach fruit development, particularly in relation to chloroplast function, photosynthesis, and response to stress. This findings from the enrichment analysis of GSE71561 align with the study conducted by Chen et al. [48], which investigated the regulation of chloroplast development in peach. The enrichment analysis highlighted gene-enriched pathways associated with cellular anatomical entity, intracellular processes, catalytic activity, membrane-bounded organelles, metabolic processes, and response to stimulus, among others. These enriched pathways provide insights into the molecular processes and organelle-specific functions involved in peach fruit development, particularly in relation to chloroplast function, photosynthesis, and response to stress.

The gene expression analysis conducted in this study focused on identifying genes associated with peach development in four gene families. The results of our analysis revealed specific genes within each family that contribute to peach development. These

findings are consistent with previous studies conducted in related species, providing valuable insights into the molecular mechanisms underlying peach development. In the Homeobox transcription factor family, we identified 8 genes associated with peach development. Previous research by Testone et al. [49] highlighted the significance of homeodomain protein Kn1 and homeobox-leucine zipper protein PROTODERMAL FACTOR 2-like in this family, specifically in stem development and regulating elongation and lignification during primary growth. These findings support our identification of Homeobox transcription factor genes associated with peach development. Similarly, within the Mad box gene family, we found 8 genes associated with peach development. In a study by Wuddineh et al. [50], a significant gene, SEEDSTICK-like protein, was identified within this family in switchgrass, a related plant species. Although their research focused on genetically engineering switchgrass for bio-fuel production, the presence of SEEDSTICK-like protein in





**Figure 4.** Gene Expression Analysis in GSE71460 and GSE71561 Datasets. (A) Heatmap illustrating the expression profiles of key genes associated with postharvest textural changes in peaches in the GSE71460 dataset, which explores the effects of gibberellic acid application and cold storage. (B) Heatmap displaying the expression patterns of critical genes related to the target trait in the GSE71561 dataset, which investigates peach fruit development, including interactions between seed and mesocarp tissues. (C) Visualization of the protein-protein interaction (PPI) network for the genes identified in the GSE71460 dataset, revealing their interactions. (D) Representation of the PPI network for the genes identified in the GSE71561 dataset, uncovering interplay among the identified genes. (E) Identification of enriched biological pathways associated with postharvest textural changes in peaches in the GSE71460 dataset, offering insights into the underlying molecular mechanisms. (F) Discovery of enriched biological pathways linked to the investigated trait in the GSE71561 dataset, providing valuable insights into fundamental molecular processes governing postharvest textural changes in peaches.

our analysis suggests its potential involvement in peach development. Within the NAC gene family, we discovered 2 genes associated with peach development, although their specific functions in relation to peach development are unknown. While no direct studies on the role of these specific genes in peach were found, further investigation of NAC genes in other plant species may provide insights into their functions in peach development. In the receptor-kinase-like gene family, we identified a total of 80 genes associated with peach development. Noteworthy genes in this family, such as LRR receptor-like serine/threonine-protein kinase RPK2 and serine/threonine-protein kinase-like protein AC-R4, have been reported in various studies across different plant species. Although no specific studies have linked these genes to peach development, their presence in our analysis suggests their potential involvement in various peach developmental processes.

### Conclusion

In conclusion, this study employed two methods, gene identification and sequence alignment, and gene expression data analysis, to investigate gene families related to *Prunus persica* (peach) and gain insights into their evolutionary relationships and regulatory mechanisms.

The gene identification and sequence alignment method involved analyzing gene families such as Homeobox, MADS-box, NAC, and Receptor kinase-like. Gene identification was conducted using local BLASTp analysis, and multiple sequence alignment was performed to assess sequence variation and construct phylogenetic trees. The results provided valuable information about gene counts, sequence lengths, and sequence variations within each gene family, shedding light on their potential roles and functions in the peach genome. The comparison with previous studies highlighted variations in gene counts and sequence lengths among different species or datasets, suggesting differences in gene composition and abundance. The gene expression data analysis method focused on analyzing two datasets, GSE71561 and GSE71460, using the GEO2R online tool. The analysis provided insights into gene expression patterns and regulatory mechanisms underlying peach fruit development, ripening, and postharvest changes. Gene enrichment analysis revealed differentially expressed genes and identified biological processes and molecular functions associated with peach fruit development and response to external stimuli such as cold storage and gibberellic acid application.

The findings from both methods contribute to a better understanding of the molecular processes governing complex phenotypes in peaches. The gene identification and sequence alignment method elucidated the diversity and potential functional divergence within gene families, providing insights into their evolutionary dynamics and functional specialization. The gene expression data analysis method revealed gene expression patterns and regulatory mechanisms underlying important developmental processes in peach.

### Reference

1. Meisel L, Fonseca B, González S, Baeza-Yates R, Cambi-azo V, Campos R, et al. A rapid and efficient method for purifying high quality total RNA from peaches (*Prunus persica*) for functional genomics analyses. *Biological research*. 2005;38(1):83-8.
2. Durst RW, Weaver GW. Nutritional content of fresh and canned peaches. *Journal of the Science of Food and Agriculture*. 2013;93(3):593-603.
3. El-Kahlout MI, Abu-Naser SS. Peach type classification using deep learning. *International Journal of Academic Engineering Research (IJAER)*. 2020;3(12).
4. Iglesias I, Echeverria G. Current situation, trends and challenges for efficient and sustainable peach production. *Scientia Horticulturae*. 2022;296:110899.
5. Liu L, Zheng C, Kuang B, Wei L, Yan L, Wang T. Receptor-like kinase RUPO interacts with potassium transporters to regulate pollen tube growth and integrity in rice. *PLoS genetics*. 2016;12(7):e1006085.
6. Preston JC, Hileman LC. Functional evolution in the plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) gene family. *Frontiers in plant science*. 2013;4:80.
7. Kaufmann K, Wellmer F, Muiño JM, Ferrier T, Wuest SE, Kumar V, et al. Orchestration of floral initiation by APETALA1. *science*. 2010;328(5974):85-9.
8. Puranik S, Sahu PP, Srivastava PS, Prasad M. NAC proteins: regulation and role in stress tolerance. *Trends in plant science*. 2012;17(6):369-81.
9. Zhong R, Demura T, Ye ZH. SND1, a NAC domain transcription factor, is a key regulator of secondary wall synthesis in fibers of Arabidopsis. *The Plant Cell*. 2006;18(11):3158-70.
10. Becraft PW, Stinard PS, McCarty DR. CRINKLY4: a TNFR-like receptor kinase involved in maize epidermal differentiation. *Science*. 1996;273(5280):1406-9.
11. Cucinotta M, Manrique S, Guazzotti A, Quadrelli NE, Mendes MA, Benkova E, et al. Cytokinin response factors integrate auxin and cytokinin pathways for female reproductive organ development. *Development*. 2016;143(23):4419-24.
12. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nature Reviews Genetics*. 2017;18(2):71-86.

13. Becker A, Theißen G. The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular phylogenetics and evolution*. 2003;29(3):464-89.
14. Diao D, Hu X, Guan D, Wang W, Yang H, Liu Y. Genome-wide identification of the ARF (auxin response factor) gene family in peach and their expression analysis. *Molecular Biology Reports*. 2020;47:4331-44.
15. Arús P, Verde I, Sosinski B, Zhebentyayeva T, Abbott AG. The peach genome. *Tree Genetics & Genomes*. 2012;8:531-47.
16. Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, et al. The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic acids research*. 2007;36(suppl\_1):D1009-14.
17. Mahram A, Herbordt MC. NCBI BLASTP on high-performance reconfigurable computing systems. *ACM Transactions on Reconfigurable Technology and Systems (TRETS)*. 2015;7(4):1-20.
18. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Current protocols in bioinformatics*. 2003;(1):2-3.
19. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic acids research*. 2019;47(W1):W256-9.
20. Alsamman AM, El Allali A, Mokhtar MM, Al-Shamaa K, Nassar AE, Mousa KH, et al. AlignStatPlot: An R package and online tool for robust sequence alignment statistics and innovative visualization of big data. *PloS one*. 2023;18(9):e0291204.
21. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data setsupdate. *Nucleic acids research*. 2012;41(D1):D991-5.
22. Li L, Tao S, Zhang H, Huang W, Dunwell JM, Li M. Identification and characterization of the CCoAOMT gene family in apple, Chinese white pear, and peach. *Journal of the American Society for Horticultural Science*. 2021;146(3):184-95.
23. Pegoraro C, Tadiello A, Girardi CL, Chaves FC, Quecini V, Costa de Oliveira A, et al. Transcriptional regulatory networks controlling woolliness in peach in response to preharvest gibberellin application and cold storage. *BMC Plant Biology*. 2015;15:1-14.
24. Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic acids research*. 2003;31(1):258-61.
25. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*. 2011;27(3):431-2.
26. Pei J, Grishin NV. AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*. 2001;17(8):700-12.
27. Chowdhury B, Garai G. A review on multiple sequence alignment from the perspective of genetic algorithm. *Genomics*. 2017;109(5-6):419-31.
28. Li Y, Zhu Y, Yao J, Zhang S, Wang L, Guo C, et al. Genome-wide identification and expression analyses of the homeobox transcription factor family during ovule development in seedless and seeded grapes. *Scientific Reports*. 2017;7(1):12638.
29. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC genomics*. 2007;8(1):1-21.
30. Fang Y, You J, Xie K, Xie W, Xiong L. Systematic sequence analysis and identification of tissue-specific or stress-responsive genes of NAC transcription factor family in rice. *Molecular Genetics and Genomics*. 2008;280:547-63.
31. Ohashi-Ito K, Fukuda H. HD-Zip III homeobox genes that include a novel member, ZeHB-13 (Zinnia)/ATHB-15 (Arabidopsis), are involved in procambium and xylem cell differentiation. *Plant and Cell Physiology*. 2003;44(12):1350-8.
32. Hirakawa Y, Kondo Y, Fukuda H. TDIF peptide signaling regulates vascular stem cell proliferation via the WOX4 homeobox gene in Arabidopsis. *The Plant Cell*. 2010;22(8):2618-29.
33. Xu X, Lou Y, Yang K, Shan X, Zhu C, Gao Z. Identification of homeobox genes associated with lignification and their expression patterns in bamboo shoots. *Biomolecules*. 2019;9(12):862.
34. Liu Y, Cui S, Wu F, Yan S, Lin X, Du X, et al. Functional conservation of MIKC\*-Type MADS box genes in Arabidopsis and rice pollen maturation. *The Plant Cell*. 2013;25(4):1288-303.
35. Alvarez-Buylla ER, Liljegen SJ, Pelaz S, Gold SE, Burgeff C, Ditta GS, et al. MADS-box gene evolution beyond flowers: expression in pollen, endosperm, guard cells, roots and trichomes. *The Plant Journal*. 2000;24(4):457-66.



36. Hu J, Chang X, Zhang Y, Yu X, Qin Y, Sun Y, et al. The pineapple MADS-box gene family and the evolution of early monocot flower. *Scientific Reports*. 2021;11(1):849.
37. Endo H, Yamaguchi M, Tamura T, Nakano Y, Nishikubo N, Yoneda A, et al. Multiple classes of transcription factors regulate the expression of VASCULAR-RELATED NAC-DOMAIN7, a master switch of xylem vessel differentiation. *Plant and Cell Physiology*. 2015;56(2):242-54.
38. Yamaguchi M, Ohtani M, Mitsuda N, Kubo M, Ohme-Takagi M, Fukuda H, et al. VND-INTERACTING2, a NAC domain transcription factor, negatively regulates xylem vessel formation in Arabidopsis. *The Plant Cell*. 2010;22(4):1249-63.
39. Zhao C, Lasses T, Bako L, Kong D, Zhao B, Chanda B, et al. XYLEM NAC DOMAIN1, an angiosperm NAC transcription factor, inhibits xylem differentiation through conserved motifs that interact with RETINOBLASTOMA-RELATED. *New Phytologist*. 2017;216(1):76-89.
40. Hendelman A, Stav R, Zemach H, Arazi T. The tomato NAC transcription factor SINAM2 is involved in flower-boundary morphogenesis. *Journal of Experimental Botany*. 2013;64(18):5497-507.
41. Takeda S, Hanano K, Kariya A, Shimizu S, Zhao L, Matsui M, et al. CUP-SHAPED COTYLEDON1 transcription factor activates the expression of LSH4 and LSH3, two members of the ALOG gene family, in shoot organ boundary cells. *The Plant Journal*. 2011;66(6):1066-77.
42. Xiang Y, Cao Y, Xu C, Li X, Wang S. Xa3, conferring resistance for rice bacterial blight and encoding a receptor kinase-like protein, is the same as Xa26. *Theoretical and Applied Genetics*. 2006;113:1347-55.
43. Zhang X, Wang L, Yuan Y, Tian D, Yang S. Rapid copy number expansion and recent recruitment of domains in S-receptor kinase-like genes contribute to the origin of self-incompatibility. *The FEBS Journal*. 2011;278(22):4323-37.
44. Saintenac C, Lee WS, Cambon F, Rudd JJ, King RC, Marande W, et al. Wheat receptor-kinase-like protein Stb6 controls gene-for-gene resistance to fungal pathogen *Zymoseptoria tritici*. *Nature genetics*. 2018;50(3):368-74.
45. Zhao D, Ni W, Feng B, Han T, Petrusek MG, Ma H. Members of the Arabidopsis-SKP1-like gene family exhibit a variety of expression patterns and may play diverse roles in Arabidopsis. *Plant physiology*. 2003;133(1):203-17.
46. Zhao L, Huang J, Zhao Z, Li Q, Sims TL, Xue Y. The Skp1-like protein SSK1 is required for cross-pollen compatibility in S-RNase-based self-incompatibility. *The Plant Journal*. 2010;62(1):52-63.
47. Kaur C, Sharma S, Singla-Pareek SL, Sopory SK. Methylglyoxal, triose phosphate isomerase, and glyoxalase pathway: implications in abiotic stress and signaling in plants. *Elucidation of Abiotic Stress Signaling in Plants: Functional Genomics Perspectives, Volume 1*. 2015:347-66.
48. Chen M, Liu X, Jiang S, Wen B, Yang C, Xiao W, et al. Transcriptomic and functional analyses reveal that PpGLK1 regulates chloroplast development in peach (*Prunus persica*). *Frontiers in Plant Science*. 2018;9:34.
49. Testone G, Condello E, Verde I, Nicolodi C, Caboni E, Dettoni MT, et al. The peach (*Prunus persica* L. Batsch) genome harbours 10 KNOX genes, which are differentially expressed in stem development, and the class 1 KNOPE1 regulates elongation and lignification during primary growth. *Journal of experimental botany*. 2012;63(15):5417-35.
50. Wuddineh WA, Mazarei M, Zhang JY, Turner GB, Sykes RW, Decker SR, et al. Identification and overexpression of a Knotted1-like transcription factor in switchgrass (*Panicum virgatum* L.) for lignocellulosic feedstock improvement. *Frontiers in plant science*. 2016;7:520.