

Genome-wide identification of UGT gene family and their expression profiles analysis during fruit development in walnut (*Juglans regia* L.)

Danhua Shi

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China

Pei Yao

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China

Jinyu Yang

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China

Gengyang Li

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China

Yanyu Shi

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China; Shanxi Provincial Key Laboratory of Fruit Tree Germplasm Creation and Utilization, Taiyuan 030031, China

Jieyun Tian

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China; Shanxi Provincial Key Laboratory of Fruit Tree Germplasm Creation and Utilization, Taiyuan 030031, China

Xiaojun Zhang

zhangx.j0906@126.com

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China; Shanxi Provincial Key Laboratory of Fruit Tree Germplasm Creation and Utilization, Taiyuan 030031, China

Qunlong Liu

lq10288@126.com

College of Horticulture, Shanxi Agricultural University, Jinzhong 030801, China; Shanxi Provincial Key Laboratory of Fruit Tree Germplasm Creation and Utilization, Taiyuan 030031, China

Research Article

Keywords: Walnut, UDP-glycosyltransferase, Gene family, Gene expression

Posted Date: April 24th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4288373/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Walnut (*Juglans regia* L.) possess the ability to prevent coronary heart disease and promote cardiovascular health, which can be attributed to their rich content of polyphenols, particularly flavonoids. The biosynthesis of flavonoids is reliant on the catalytic activity of uridine diphosphate glycosyltransferase (*UGT*). However, the identification of *UGTs* in walnut has not been reported. In the current study, a total of 124 *UGT* genes containing the PSPG box were identified from the walnut genome. Based on phylogenetic analysis, the 124 *UGT* could be classified into 16 distinct groups, which exhibited an uneven distribution across the 16 chromosomes. Subcellular localization prediction analysis revealed that approximately 78.23% of walnut *UGT* proteins were predominantly localized in the cytoplasmic compartment. Furthermore, motif annotation confirmed that motifs 1, 2, and 3 represented conserved structural features within *UGT* proteins, while interestingly, around 56.5% of walnut *UGT* members lacked introns. In conjunction with transcriptome analysis and quantitative expression, approximately 39% of *UGT* genes in walnut exhibited high expression levels during early fruit development. Correlation analysis between *UGT* genes expression and phenolic content in walnut indicated that *JrUGT6*, *JrUGT38*, *JrUGT39*, *JrUGT58*, *JrUGT69*, *JrUGT75*, and *JrUGT82* might be involved in phenolic biosynthesis in walnut. This comprehensive study provides a comprehensive overview of the *UGT* genes in walnut, serving as a valuable reference and theoretical foundation for further investigations into the biological functions of *JrUGTs* in flavonoid biosynthesis.

Introduction

Walnut (*Juglans regia* L.) is a member of the Juglans family and extensively cultivated worldwide. Many evidence suggested that walnuts possess the ability to prevent coronary heart disease and promote cardiovascular health, which can be attributed to their rich content of polyphenols, particularly flavonoids (Feldman et al. 2002; Mc et al. 2012). Flavonoids, as secondary metabolites in plants, represent a class of polyphenolic compounds that exhibit potent biological activity and are widely distributed throughout the plant kingdom. The biosynthesis of flavonoids relies on the catalytic function of uridine diphosphate glycosyltransferase family (*UGT*), especially Uridinediphosphate Glycosyltransferase (Shirley et al. 2001). Glycosylation, an important modification reaction in plants, plays a crucial role in growth and response to biotic and abiotic stresses, and it also enhances solubility, stability, transferability, and diversity of various plant secondary metabolites such as flavonoids (Vogt et al. 2000; Bowles et al. 2006; Yang et al. 2018; Naeem et al. 2021).

Glycosylation modification is catalyzed by glycosyltransferases (GTs), which are highly diverse, multifunctional enzymes belonging to a polymorphic gene family (Mackenzie et al. 1997). According to the latest update on CAZy (<http://www.cazy.org/GlycosylTransferases.html>), GTs can be classified into 115 distinct species based on amino acid sequence similarity, catalytic mechanism, and the presence of conserved sequence motifs (Chen et al. 2022). Among these, the GT1 family, commonly referred to as uridine diphosphate glycosyltransferases (*UGTs*), represents the predominant group in plants and exerts significant influence on plant growth and development (Yonekura et al. 2011). The *UGTs* not only glycosylate receptor molecules, such as anthocyanins, flavonols, flavonoids, saponins, sterol terpenes, phenylpropanes and plant hormones but also play a crucial role in the detoxification and inactivation of foreign organisms and plant-pathogen interactions (Vogt et al. 2000; Caputi et al. 2012). The *UGT* protein exhibits a highly conserved sequence of 44 amino acids (aa) near the C-terminal region and is referred to as the plant secondary product glycosyl transferase (PSPG) (Zhang et al. 2020). Moreover, *UGT* demonstrates broad substrate specificity in plants by recognizing multiple receptor molecules selectively (Taguchi et al. 2000; Lim et al. 2003).

A growing number of putative *UGT*-coding genes have been discovered in plants, with 107 members identified in *Arabidopsis thaliana*, which were categorized into 14 groups based on their amino acid sequences and designated as groups A-N (Li et al. 2001). Subsequently, *Zea mays* was found to possess 147 *UGT* genes, *Vitis vinifera* had 181, *Malus x domestica* contained 241, *Citrus grandis* 145 and *Camellia sinensis* harbored 178. (Velasco et al. 2010; Li et al. 2014; Cui et al. 2016; Wu et al. 2020). Polygenic *UGT* family members collaborate synergistically to orchestrate intricate biochemical processes in plant cells, thereby exerting profound influence on diverse biological activities and functions. Currently, the functionality of *UGT* genes has been validated in multiple species. Yang et al (2023) demonstrated the functions of 276 glycosyltransferases in *Nicotiana tabacum* and observed a significant increase in the levels of quercetin-3-*O*-glucoside, quercetin-3-*O*-rutin, and kaempferol-3-*O*-rutin in transgenic tobacco leaves by overexpressing *NtUGT217*. In a study by Lim et al (2004), 91 *UGT* genes were identified in *Arabidopsis thaliana* using quercetin as a substrate, with 29 of them exhibiting catalytic activity towards related glycosylation reactions. Yao et al (2022) conducted a comprehensive genome-wide analysis of epimedium, revealing that only the recombinant EpGT60 protein exhibited activity against 8-enyl kaempferol and icariin, leading to the biosynthesis of key bioactive compounds pyridoside II and pyridoside I. However, despite the large number of *UGT* genes in the plant genome, the functional validation of signature proteins remains limited (Caputi et al. 2012).

The present study employed whole genome sequencing to identify the *UGT* gene family in walnut (*JrUGT*). Subsequently, sequence alignment and phylogenetic tree analysis were conducted to elucidate the genetic relationships among these *UGTs* in walnut. Furthermore, predictions regarding the physical and chemical properties of proteins, subcellular localization, chromosome distribution, collinearity, and gene structure were made for *JrUGT*. Based on this premise, the expression patterns of *UGT* genes in various developmental stages of walnut kernel pellicle were analyzed and subsequently validated using qRT-PCR. The correlation analysis between *JrUGT* and phenolic content establishes a fundamental for future investigations into the biosynthesis mechanism and accumulation pattern of walnut phenolics.

Materials and methods

Identification of UGT gene family in walnut

The genome sequence information of walnut was primarily obtained from the *Juglans* genome database (<http://www.juglandaceae.net/>). *Arabidopsis* *UGT* members were predominantly retained from the TAIR database (<https://www.arabidopsis.org/>). The Hidden Markov Model HMM file for the conserved domain of *UGT* gene family (PF00201) was accessible through Pfam (<http://pfam.xfam.org/>). By employing the *Arabidopsis* *UGT* amino acid sequence as a query, BLASTP was utilized to search for proteins in the walnut genome with an e-value threshold set at $1e^{-15}$, thereby constructing *UGT* candidate dataset 1. The hmmsearch function in TBtools was employed to retrieve the *UGT* candidate data 2 by utilizing the HMM file of the *UGT* family as a probe. Subsequently, two sets of candidate data were combined and subjected to verification using the NCBI CDD program, confirming the presence of conserved domain PSPG box in all walnut *UGT* members.

Analysis of physicochemical properties of UGT gene family

The Expasy online database (<http://www.expasy.org/>) was utilized for the prediction of physicochemical properties of the amino acid sequence, encompassing relative molecular weight, theoretical isoelectric point, and amino acid length. Additionally, subcellular localization analysis of walnut *UGT* protein was performed using the PSORT II Prediction online software (<https://psort.hgc.jp/form2.html>).

Phylogenetic and chromosomal localization analysis

The *UGT* protein sequences of walnut, *Arabidopsis*, *Zea mays* and *Pilosella officinarum* were aligned with MUSCLE software. Subsequently, phylogenetic trees based on 1000 bootstrap replicates using the neighbor-joining (NJ) method were constructed in MEGA 11 software (Kumar et al. 2016). The genomic localization information of *JrUGT* members was visually analyzed using TBtools software, as reported by the walnut genome database.

Gene replication and collinearity analysis

The gene replication events were analyzed using the One Step MCScanX program and visualized with the TBtools program (Wang et al. 2012; Chen et al. 2020). To investigate the evolutionary relationship between walnut and *Arabidopsis thaliana*, the gene sequence of walnut was comprehensively analyzed (Chen et al. 2020).

Conserved sequence and gene structure analysis

The conserved motifs of walnut *UGT* members were identified using the MEME program (<http://meme-suite.org/tools/meme>). The conserved motifs recognized by MEME programs were further annotated using the NCBI CDD program, with a total of 10 motifs identified. Additionally, the walnut gene annotation files and coding sequences from its corresponding genome sequence comparison were inputted into the NCBI GSDS online tools (v2.0) at <http://gsds.cbi.pku.edu.cn/> to display information on exons and introns of walnut *UGT*.

Expression pattern analysis of 13 JrUGTs

The kernel pellicle of the walnut varieties 'NongHe 1' ('NH1') and 'FenHe 4' ('FH4') were collected at different developmental stages, including the hard core stage, the fatty stage and the mature stage. These samples were obtained from the walnut planting resource nursery of the horticultural experimental station of Shanxi Agricultural University. Subsequently, all samples were rapidly frozen in liquid nitrogen and stored at -80°C in an ultra-low temperature refrigerator for future use. Three biological replicates were performed on each

sample. The kernel pellicle of two walnut varieties, 'NH1' and 'FH4', at various developmental stages were submitted to Baimaike Biotechnology Co., Ltd. for transcriptome sequencing analysis. Transcript abundance was quantified using fragments per kilobase transcript per million fragment mapping (FPKM) values derived from RNA-seq reads. Each gene exhibited distinct expression levels. The tissue-level data was averaged and presented as $\log^2$ values, followed by the generation of a heat map using TBTOOLS software (Anders et al. 2010).

The total RNA was extracted using the improved CTAB method (Feng et al., 2013), and the first strand cDNA synthesis was performed using the HiScript II 1st Strand cDNA Synthesis Kit (+ gDNA wiper). Specific primers for 13 selected *JrUGTs* were designed using Primer Premier 5.0 software, and their sequences were provided in Table S1. The relative expression of *JrUGTs* was determined by qRT-PCR reaction system with the $2^{-\Delta\Delta CT}$ method (Livak et al. 2001).

Determination of phenolics

The kernel pellicle was freeze-dried and ground into a fine powder. 0.1 g of the kernel pellicle powder was measured and added to 5 mL of a 50% methanol solution. Ultrasound-assisted extraction was conducted for 30 min at 50°C, followed by centrifugation at 12000 r/min for 10 min to obtain the supernatant. This process was repeated twice, and the resulting supernatants were combined and adjusted to a constant volume before being analyzed using HPLC. The content of phenolic compounds was determined following the methods described by Lu et al (2017) and Yang et al (2020). The regression equation for the standard curve is presented (Table S2).

Results

Identification of UGT gene family members in walnut

In order to identify *JrUGT*, *UGT* proteins in walnut genome database were screened by using HMM file (PF00201) as a query. The screening criterion was set at an e-value < 1. We employed two screening methods, namely BLASTP and hmmsearch, resulting in the identification of 174 potential *UGT* genes in walnut. To verify the presence of conserved domains, TBtools was employed, and any *UGT* proteins lacking a PSPG box were manually excluded. Ultimately, 124 *JrUGT* genes were obtained, and designated as *JrUGT1-124*.

The 124 *JrUGTs* genes encoded proteins with diverse physicochemical properties (Table S3). Each member of the *JrUGT* family had a protein length ranging from 216 to 611 aa, with an average length of 474 aa. The theoretical molecular weight of the protein ranges from 29.29 to 61.50 KDa, with an average value of 52.90 KDa. The theoretical isoelectric point (pI) ranges from 5.17 to 8.1, with an average pI of 5.85. According to the PSORT II Prediction online tool for subcellular localization prediction, majority of the *UGT* proteins in walnut (78.23%) were localized in the cytoplasm, followed by mitochondria (12.10%). Additionally, eleven *UGT* proteins were found in the endoplasmic reticulum and one protein was located extracellularly within the cell wall.

Phylogenetic and chromosomal localization analysis

The classification and phylogenetic relationship of *UGT* proteins in walnut were investigated by constructing a phylogenetic tree based on *UGT* protein sequences from walnut, *Arabidopsis thaliana*, *Zea mays*, and *Pilosella officinarum*. The main clusters of walnut *UGT* family members consisted of 18 subgroups, with a total of 124 *JrUGTs* being classified into these previously identified subgroups (Fig. 1). Notably, *JrUGTs* were absent from groups F and Q, while the majority of them clustered within groups E (25), G (16), D (19), L (15), and A (15). It has been reported that *AtUGT71B1*, *AtUGT72B1*, and *AtUGT88A1* of *Arabidopsis* were involved in flavonoid biosynthesis. Sequence analysis found *JrUGT4*, *JrUGT13*, *JrUGT31*, *JrUGT32*, *JrUGT43*, *JrUGT44*, *JrUGT45*, *JrUGT57*, *JrUGT58*, *JrUGT59*, *JrUGT60*, *JrUGT61*, *JrUGT62*, *JrUGT67*, *JrUGT68*, *JrUGT69*, *JrUGT80*, *JrUGT81*, *JrUGT93*, *JrUGT108*, *JrUGT111*, *JrUGT114*, *JrUGT117*, *JrUGT118*, *JrUGT121* clustered with *AtUGT71B1*, *AtUGT72B1*, and *AtUGT88A1*, suggesting that these *JrUGT* may be involved in flavonoid biosynthesis in walnut.

Among the 124 *JrUGTs* identified, 124 *JrUGTs* were located on 16 chromosomes of walnut (Fig. 2). Among the 16 walnut chromosomes, chromosome 9 contained only one *UGT* family member, while chromosome 1 contains 17 *UGT* genes. Chromosome 7 contained 15 *UGT* genes, chromosome 3 contained 13 *UGT* genes, and chromosomes 2 and 15 each contained only 2 *UGT* genes. This unbalanced distribution of *UGT* gene in walnut on chromosome indicates that there was genetic variation in walnut during evolution.

Gene replication and collinearity analysis

Gene replication events play a pivotal role in the formation of gene families. In order to elucidate the expansion and evolution mechanism of *UGT* gene family in walnut, the potential gene replication events in walnut genome were further investigated. The

detection of *UGT* gene replication events in walnut was performed using Tbttools software. A total of five genes located on chromosomes 1, 2, 4, 9, 10, and 14 were identified to undergo four gene replication events (Fig. 3). Notably, chromosomes 1 and 10 exhibited the highest frequency of tandem repeat events (three occurrences each), implying that *UGT* genes might have originated from such replication events which could be considered as key drivers for *UGT* evolution. Furthermore, by comparing the DNA sequence similarity between the *UGT* gene of walnut and the homologous genes of *Arabidopsis thaliana* (Fig. 4), a collinear relationship between 39 walnut genes and 36 *Arabidopsis* genes was found. These conserved genes are likely to possess crucial functions across different species.

Conserved sequence and gene structure analysis

To further elucidate the conserved domain characteristics of the walnut *UGT* family, 10 motifs were generated using the online tool MEME and then numbered from 1 to 10. Notably, motif 1 and motif 3 corresponded to the highly conserved PSPG box within the *UGT* family. The distribution pattern of these motifs among different types of walnut *UGT* members was depicted (Fig. 5). Remarkably, our findings revealed that members belonging to the same group exhibited either identical or similar conserved motifs. The characteristic sequence motif 1–3, present in all walnut *UGT* proteins, was considered to be the glycosyltransferase recognition site for glycosyl-donor. With a few exceptions, most walnut *UGT* proteins exhibited the following characteristics: motif 6 was located proximal to motif 2, and motif 5 was also positioned near motif 8. In the majority of sequences, motif 4 appeared at the beginning while motif 7 occurred at the end. The protein's sequence typically follows this pattern: motif 4-5-8-6-2-10-1-3-9-7; however, variations existed among certain proteins.

The diversity in intron-exon structure often plays a pivotal role in the evolutionary dynamics of gene families and provides supplementary evidence to support phylogenetic classifications. In order to further understand the gene structure, the intron-exon structure of *UGT* gene in walnut was analyzed. Out of the 124 *UGT* genes identified in this study, 47 contained a single intron, 70 were devoid of any introns, while 7 harbored two introns.

Expression patterns of UGT genes

The transcriptomic sequencing results of the kernel pellicle of two walnut varieties, 'NH1' and 'FH4', at different developmental stages were analyzed in this study to gain further insights into the expression pattern of *UGT* genes in walnut. The findings revealed that out of 118 *UGT* genes analyzed in walnut, *JrUGT2*, *JrUGT11*, *JrUGT68*, *JrUGT113*, *JrUGT119* and *JrUGT121* were not detected in the transcriptome data. A total of 19 *JrUGTs* (FPKM > 10) exhibited high transcription levels in 'NH1' and 'FH4', respectively (Fig. S1). The expression profiles of *JrUGT39*, *JrUGT58*, *JrUGT75*, and *JrUGT118* were consistently elevated across all three developmental stages in both varieties, suggesting their potential involvement in diverse biological processes throughout growth and development. Group E, which represents the largest subset of the *UGT* gene family in walnut, exhibited predominant expression of *UGT* during the hard core stage in 'NH1' and 'FH4'. Additionally, some genes showed higher expression levels in the mature stage compared to those observed in the hard core stage and the fatty stage. Similarly, Group G also displayed peak expression during the hard core stage with greater abundance detected in 'FH4' than 'NH1'. Most *UGT* genes demonstrated elevated transcript levels during early fruit development stages but decreased as fruit matured.

In this study, we performed a comprehensive analysis of the walnut *UGT* gene family through phylogenetic tree analysis and transcriptome differential gene screening. Subsequently, 13 differentially expressed walnut *UGT* genes were randomly selected for validation using real-time fluorescence quantitative PCR. The qRT-PCR results were presented (Fig. 6). The gene expression levels of *JrUGT39*, *JrUGT58*, *JrUGT88*, and *JrUGT95* were found to be higher in August, while *JrUGT118* exhibited the lowest gene expression levels during this period. These findings were consistent with the transcriptome data analysis.

Analysis of phenolic substances in walnut kernel pellicle at different periods

HPLC was used to determine the content of 'NH1' and 'FH4' phenolics in walnut kernel pellicle at different periods. With the growth and development of walnut fruit, the content of phenolics also changed, and the results were shown (Table 1). The content of Gallic acid was the highest among the 11 substances and the content of GCG was the lowest. The content of C was the highest among catechins, and the content of C in 'FH4' was significantly higher than that of 'NH1'. The contents of EC and GC showed a trend of continuous increase on the whole, and the contents of EGC and EGCG in the two varieties showed a trend of first increasing and then decreasing, and the content of EGC in 'NH1' fatty stage was significantly higher than that in other periods. The content of 'FH4' in syringate was higher than that of 'NH1' in different periods.

Correlation analysis of UGT gene expression and phenolic content in walnut

The Pearson coefficient was employed to investigate the correlation between *UGT* gene expression and phenolic substance content in walnut, and the results were presented (Fig. 7). No significant correlation was observed between *JrUGT36* and phenolic substances. Only *JrUGT118* showed a positive correlation with EGCG and Chlorogenic acid ($P < 0.05$). *JrUGT6*, *JrUGT38*, *JrUGT39*, *JrUGT58*, *JrUGT69*, *JrUGT75*, and *JrUGT82* exhibited positive correlations with Vanillic acid ($P < 0.01$). Additionally, *JrUGT38*, *JrUGT39*, *JrUGT67*, *JrUGT69*, *JrUGT75*, and *JrUGT82* were positively correlated with C, GC, and EC. These findings suggest that these identified *UGTs* may play a role in the biosynthesis of phenolic substances in walnut; thus providing valuable insights for further study on this topic.

Discussion

Glycosylation is modified by GTs, which can be divided into at least 111 families, of which the *UGT* gene family is the largest family (Paquette et al, 2003). Currently, the *UGT* gene family has been found and analyzed in some plants such as *Arabidopsis*, *Triticum aestivum*, *Zea mays* and *Morella rubra* (Li et al, 2001; He et al, 2018; Li et al, 2014; Ren et al., 2022). However, there was a lack of information regarding the *UGT* gene family in walnut. Through BLASTP and hmsearch screenings along with conserved domain verification, 124 *JrUGTs* were identified. Compared to other species, the number of *UGT* genes in walnut showed lower numbers, that 152 *UGT* genes in *Morella rubra* and 168 *UGT* genes in *Prunus persica* (Wu et al. 2017). This observation suggests that the amplification of the walnut *UGT* gene family was not significant, possibly due to the absence of genome-wide duplication events in walnut. The protein encoded by the *UGT* gene of walnut exhibits a wide range of amino acid lengths and molecular weights. The intron content in walnut *UGT* family members was 43.5%, which was lower compared to *Arabidopsis* (58%) (Li et al. 2001), *Linum usitatissimum* (55%) (Barvkar et al. 2012), and *Zea mays* (60%) (Li et al. 2014). To functionally identify *UGT* genes in walnut, the phylogenetic tree was constructed to classify the 124 identified *UGT* genes into 16 groups. In the evolutionary process of higher plants, groups A, L, D, G, and E of the *UGT* gene family were considered as rapidly evolving groups (Caputi et al. 2012), and these five groups also contain the highest number of *UGT* gene family members in walnut, consistent with the aforementioned results. The phylogenetic tree showed that group E contained the most *UGT* genes (25), accounting for 20.2% of the total *UGT* genes in walnut. There were 17, 22 and 35 *UGT* genes in group E in *Arabidopsis*, *Linum usitatissimum* and *Zea mays*, respectively, indicating the expansion of group E in different plant species (Barvkar et al., 2012; Caputi et al., 2012). *JrUGTs* were not distributed in group F and group Q. Group F was first identified in *Arabidopsis Thaliana* (Li et al., 2001), and group Q was first identified in *Zea mays* (Li et al., 2014), and this group of genes was thought to be related to cytokinin glycosylation.

124 *JrUGTs* were located on 16 chromosomes. In *Dendrobium nobile* and *Gossypium*, the distribution of genes on chromosomes usually existed in clusters with high homology, and the distribution of *UGT* gene family on chromosomes of walnut also showed this feature (Ren et al, 2020; Huang et al., 2015). Qiao et al. (2019) argued that gene replication events include five modes, namely WGD, TD, DSD, PD and TRD. Tandem repeat events play an important role in the expansion of gene families. In *Broussonetia papyrifera*, TD was the main driver of expansion of the *BpUGT* gene family (Wang et al, 2021). In the *UGT* family of walnut, a total of 5 tandem replicators were obtained, and four tandem replication events occurred. Therefore, tandem repetition events continue to occur in *UGT* gene families, which is a continuous process throughout the evolutionary history of *UGT* gene families. By analyzing the conserved motifs of *UGT* proteins in walnut, 10 different motifs were found, and all *UGT* proteins contained motif 1–3. The *UGT* protein of walnut in group E did not contain motif 9. These specific motifs may lead to differentiation of *UGTs* function in walnut.

Understanding the spatiotemporal expression patterns of genes is crucial for predicting their functional roles. In 'NH1', a total of 62 *UGT* genes exhibited significant expression (FPKM > 1) during kernel pellicle development, while 'FH4' showed the expression of 52 *UGT* genes with FPKM > 1. Notably, both varieties displayed higher expression levels of *JrUGT58* and *JrUGT118* in group E throughout fruit development. Additionally, *JrUGT75* demonstrated the highest level of expression within group G. Interestingly, 'FH4' exhibited a tenfold increase in FPKM value for *JrUGT67* compared to 'NH1'; conversely, *JrUGT69* displayed an opposite pattern of expression between these two varieties. These specific gene expressions may potentially impact the metabolite composition in walnut fruits. Correlation analysis between *UGT* gene expression and phenolics content in walnut showed that *JrUGT36* had no significant correlation with phenolics. *JrUGT6*, *JrUGT38*, *JrUGT39*, *JrUGT58*, *JrUGT69*, *JrUGT75* and *JrUGT82* were significant correlation with C, GC, EC and Vanillic acid. This discovery improved the basis for subsequent research on phenolic biosynthesis of walnut.

Conclusion

In this study, a total of 124 *JrUGTs* protein sequences containing PSPG boxes were identified in the walnut genome, and they were categorized into 16 functional groups. The *UGT* protein motifs and gene structures in walnut exhibited similarities within the same group, while notable differences were observed between different groups. Significantly distinct expression patterns of *UGT* gene family

members during various developmental stages were detected in the kernel pellicle of 'NH1' and 'FH4'. Correlation analysis between *UGT* gene expression and phenolic content in walnut indicated that *JrUGT6*, *JrUGT38*, *JrUGT39*, *JrUGT58*, *JrUGT69*, *JrUGT75*, and *JrUGT82* might be involved in phenolic biosynthesis in walnut. The results provided a basis for further study on the biosynthesis of phenolic substances in walnut.

Declarations

Author Contributions

All authors contributed to the conception and design of the study. XZ and QL conceived and designed the experiment; DS, PY analysis data; DS and JY write papers; DS, GL and YS conducted experiments. XZ, QL, and JT reviewed the papers, and all authors read and approved the final manuscript.

Funding

This work was supported by the Basic Research Program of Shanxi Province (No. 20210302123396); the science and technology major project of Shanxi Province (No. 202201140601027) and the Key Research and Development Plan (Agriculture) of Jinzhong City (No. Y212013).

Conflict of Interest

All authors declare no conflict of interest.

References

1. Anders S, Huber W (2010) Differential expression analysis for sequence count data. *Genome Biology* 11: R106. <https://doi.org/10.1186/gb-2010-11-10-r106>
2. Barvkar V T, Pardeshi V C, Kale S M, Kadoo N Y, Gupta V S (2012) Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in *Linum usitatissimum* identified genes with varied expression patterns. *BMC genomics* 13: 1-13. <https://doi.org/10.1186/1471-2164-13-175>
3. Bowles D, Lim E K, Poppenberger B, Vaistij F E (2006) Glycosyltransferases of lipophilic small molecules. *Annual Review of Plant Biology* 57: 567-597. <https://doi.org/10.1146/annurev.arplant.57.032905.105429>
4. Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *The Plant Journal* 69(6): 1030-1042. <https://doi.org/10.1111/j.1365-313X.2011.04853.x>
5. Chen C, Chen H, Zhang Y, Thomas H R, Frank M H, He Y (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* 13:1194-1202. <https://doi.org/10.1016/j.molp.2020.06.009>
6. Chen Y, Fu M, Li H, Wang L, Liu R, Liu Z (2022) Genome-wide characterization of the UDP-glycosyltransferase gene family reveals their potential roles in leaf senescence in cotton. *International Journal of Biological Macromolecules* 222:2648–2660. <https://doi.org/10.1016/j.ijbiomac.2022.10.047>
7. Cui L L, Yao S, Dai X, Yin Q, Liu Y, Jiang X (2016) Identification of UDP-glycosyltransferases involved in the biosynthesis of astringent taste compounds in tea (*Camellia sinensis*). *Journal of experimental botany* 67(8): 2285-2297. <https://doi.org/10.1093/jxb/erw053>
8. Feldman E B (2002) The scientific evidence for a beneficial health relationship between walnuts and coronary heart disease. *The Journal of nutrition* 132(5): 1062S-1101S. <https://doi.org/10.1038/sj.jo.0801981>
9. Feng C, Xu C J, Wang Y, Liu W L, Yin X R, Li X (2013) Codon usage patterns in Chinese bayberry (*Myrica rubra*) based on RNA-Seq data. *BMC Genomics* 14:732. <https://doi.org/10.1186/1471-2164-14-732>
10. He Y, Ahmad D, Zhang X, Zhang Y, Wu L, Jiang P, Ma H (2018) Genome-wide analysis of family-1 UDP glycosyltransferases (UGT) and identification of UGT genes for FHB resistance in wheat (*Triticum aestivum* L.). *BMC plant biology* 18(1): 67. <https://doi.org/10.1186/s12870-018-1286-5>
11. Huang J, Pang C, Fan S, Song M, Yu J, Wei H, Yu S (2015) Genome-wide analysis of the family 1 glycosyltransferases in cotton. *Molecular Genetics and Genomics* 290: 1805-1818. <https://doi.org/10.1007/s00438-015-1040-8>

12. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution* 33(7): 1870-1874. <https://doi.org/10.1093/molbev/msw054>
13. Lim E K, Ashford D A, Hou B, Jackson R G, Bowles D J (2004) *Arabidopsis* glycosyltransferases as biocatalysts in fermentation for regioselective synthesis of diverse quercetin glucosides. *Biotechnology and bioengineering* 87(5): 623-631. <https://doi.org/10.1002/bit.20154>
14. Lim E K, Baldauf S, Li Y, Elias L, Worrall D, Spencer S P, Bowles D J (2003) Evolution of substrate recognition across a multigene family of glycosyltransferases in *Arabidopsis*. *Glycobiology* 13(3): 139-145. <https://doi.org/10.1093/glycob/cwg017>
15. Livak K J, Schmittgen T D (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* 25(4): 402-408. <https://doi.org/10.1006/meth.2001.1262>
16. Li Y, Baldauf S, Lim E K, Bowles D J (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *Journal of Biological Chemistry* 276: 4338-4343. <https://doi.org/10.1074/jbc.M007447200>
17. Li Y J, Li P, Wang Y, Dong R, Yu H, Hou B (2014) Genome-wide identification and phylogenetic analysis of Family-1 UDP glycosyltransferases in maize (*Zea mays*). *Planta* 239: 1265-1279. <https://doi.org/10.1007/s00425-014-2050-1>
18. Lu W Q, Zhou X M, Zhang H Y, Jiang J S, Fu L (2017) Simultaneous determination of 9 polyphenols in fruits by high performance liquid chromatography. *Chinese Journal of Health Laboratory Technology* 2(27):317-320. <https://doi.org/CNKI:SUN:ZWJZ.0.2017-03-005>
19. Mackenzie P I, Owens I S, Burchell B, Bock K W, Bairoch A, Belanger A, Nebert D W (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics and Genomics* 7(4): 255-269. <https://doi.org/10.1097/00008571-199708000-00001>
20. McGranahan G, Leslie C. Walnut[J]. *Fruit breeding* 2012: 827-846. doi:10.1007/978-14419-0763-9_22
21. Naeem A, Yang M, Hu P Y, Kang Y J, Ya L L, Hai Y Z, (2021) The fate of flavonoids after oral administration: A comprehensive overview of its bioavailability. *Critical Reviews in Food* 4:1-18. <https://doi.org/10.1080/10408398.2021.1898333>.
22. Paquette S, Møller B L, Bak S (2003) On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62(3): 399-413. [https://doi.org/10.1016/S0031-9422\(02\)00558-7](https://doi.org/10.1016/S0031-9422(02)00558-7)
23. Qiao X, Li Q, Yin H, Qi K, Li L, Wang R, (2019) Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biology* 20:1-23. <https://doi.org/10.1186/s13059-019-1650-2>
24. Ren C, Cao Y, Xing M, Guo Y, Li J, Xue L, Sun C, Xu C, Chen K and Li X (2022) Genome-wide analysis of UDP-glycosyltransferase gene family and identification of members involved in flavonoid glucosylation in Chinese bayberry (*Morella rubra*). *Front Plant Sci* 13:998985. <https://doi.org/10.3389/fpls.2022.998985>
25. Ren Z, Ji X, Jiao Z, Luo Y, Zhang G Q, Tao S, Wei G (2020) Functional analysis of a novel C-glycosyltransferase in the orchid *Dendrobium catenatum*. *Horticulture Research* 7:111. <https://doi.org/10.1038/s41438-020-0330-4>
26. Taguchi G, Imura H, Maeda Y, Kodaira R, Hayashida N, Shimosaka M, Okazaki M (2000) Purification and characterization of UDP-glucose: hydroxycoumarin 7-O-glucosyltransferase, with broad substrate specificity from tobacco cultured cells. *Plant Science* 157(1): 105-112. [https://doi.org/10.1016/S0168-9452\(00\)00270-3](https://doi.org/10.1016/S0168-9452(00)00270-3)
27. Velasco R, Zharkikh A, Affourtit J, Dhingra A, Cestaro A, Kalyanaraman A, Viola R (2010) The genome of the domesticated apple (*Malus domestica* Borkh.). *Nature Genetics* 42(10): 833-839. <https://doi.org/10.1038/ng.654>
28. Vogt T, Jones P (2000) Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. *Trends in plant science* 5(9): 380-386. [https://doi.org/10.1016/S1360-1385\(00\)01720-9](https://doi.org/10.1016/S1360-1385(00)01720-9)
29. Wang Y, Tang H, Debarry J D, Tan X, Li J, Wang X, et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research* 40: e49. <https://doi.org/10.1093/nar/gkr1293>
30. Wang F, Su Y, Chen N, Shen S (2021) Genome-wide analysis of the UGT gene family and identification of flavonoids in *Broussonetia papyrifera*. *Molecules* 26(11): 3449. <https://doi.org/10.3390/molecules26113449>
31. Winkel-Shirley B (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology* 126(2): 485-493. <https://doi.org/10.1104/pp.126.2.485>
32. Wu B, Gao L, Gao J, Xu Y, Liu H, Cao X (2017) Genome-wide identification, expression patterns, and functional analysis of UDP glycosyltransferase family in peach (*Prunus persica* L. Batsch). *Frontiers In Plant Science* 8: 251763. <https://doi.org/10.3389/fpls.2017.00389>

33. Wu B, Liu X, Xu K, Zhang B (2020) Genome-wide characterization, evolution and expression profiling of UDP-glycosyltransferase family in pomelo (*Citrus grandis*) fruit. BMC Plant Biology 20:1–12. <https://doi.org/10.1186/S12870-020-02655-2>
34. Yang B, Liu H, Yang J, Gupta V K, Jiang Y (2018) New insights on bioactivities and biosynthesis of flavonoid glycosides. Trends In Food Science & Technology 79: 116-124. <https://doi.org/10.1016/j.tifs.2018.07.006>
35. Yang J C, Bai X M (2020) Simultaneous Determination of Catechins and Caffeine Content in Tea by HPLC. Guizhou Agricultural Sciences 48(2): 99-102. <https://doi.org/CNKI:SUN:GATE.0.2020-02-021>
36. Yang Q, Zhang Y, Qu X, Wu F, Li X, Ren M, Chen S (2023) Genome-wide analysis of UDP-glycosyltransferases family and identification of UGT genes involved in abiotic stress and flavonol biosynthesis in *Nicotiana tabacum*. BMC Plant Biology 23(1): 204. <https://doi.org/10.1186/s12870-023-04208-9>
37. Yao Y, Gu J, Luo Y, Wang Y, Pang Y, Shen G, Guo B (2022) Genome-wide analysis of UGT gene family identified key gene for the biosynthesis of bioactive flavonol glycosides in *Epimedium pubescens* Maxim. Synthetic and Systems Biotechnology 7(4): 1095-1107. <https://doi.org/10.1016/j.synbio.2022.07.003>
38. Yonekura-Sakakibara K, Hanada K (2011) An evolutionary view of functional diversity in family 1 glycosyltransferases. The Plant Journal 66:182–193. <https://doi.org/10.1111/j.1365-313X.2011.04493.x>
39. Zhang P, Zhang Z, Zhang L, Wang J, Wu C (2020) Glycosyltransferase GT1 family: Phylogenetic distribution, substrates coverage, and representative structural features. Computational and Structural Biotechnology Journal 18:1383–1390. <https://doi.org/10.1016/j.csbj.2020.06.003>

Tables

Table 1 Content analysis of phenolics in different stages (mg/100g). Different lowercase letters in the same line indicate significant difference in kernel pellicle at 0.05 level in different periods.

Developmental stage	NH1-6	NH1-7	NH1-8	FH4-6	FH4-7	FH4-8
Substance name						
C	866.71±2.71b	822.45±16.30c	746.36±12.40e	733.57±5.98e	785.20±11.82d	1797.77±13.64a
GC	168.22±2.31e	203.57±2.49c	355.09±2.61b	144.95±4.70f	185.72±4.39d	591.43±4.24a
EC	56.97±4.37e	230.58±2.91c	213.56±2.17d	57.44±2.91e	243.31±6.38b	315.93±2.13a
EGCG	209.25±6.12bc	254.38±28.44a	152.46±10.50d	231.61±1.67ab	232.77±4.42ab	191.65±2.40c
EGC	318.98±7.78d	646.48±13.98a	186.41±12.01f	282.21±15.76e	369.06±7.51b	344.71±4.98c
GCG	39.93±4.05b	8.18±0.40e	15.21±0.30cd	13.07±1.69d	17.52±0.58c	45.02±2.73a
ECG	359.08±25.51a	296.73±24.67b	177.76±4.80c	151.43±0.80cd	154.68±7.26cd	145.33±1.43d
Gallic acid	1390.92±46.78b	1535.90±11.41a	1223.70±4.08c	1086.18±11.21d	1243.39±19.68c	1218.43±15.23c
Chlorogenic acid	366.99±9.65c	263.89±7.50d	142.40±4.66e	547.09±32.01a	256.41±6.06d	449.65±11.78b
vanillic acid	185.66±5.70f	333.59±6.35c	321.29±2.72d	230.55±3.22e	451.43±11.37b	645.98±2.52a
Syringate	55.07±3.76e	97.62±7.02c	63.65±3.50e	76.84±6.60d	186.88±9.21a	139.95±1.75b

Figures

Fig.1

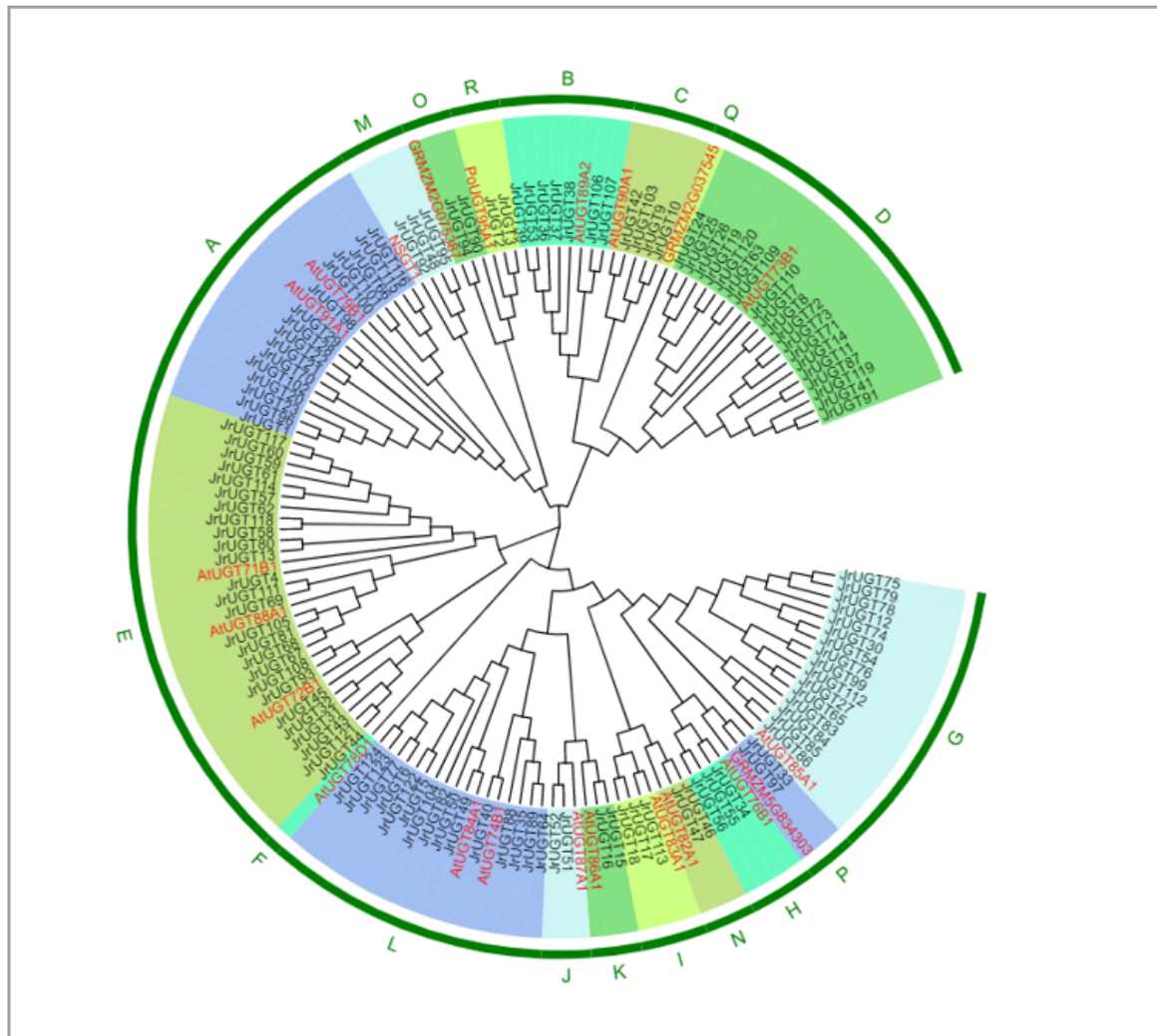


Figure 1

Phylogenetic tree was constructed with *UGT* proteins, encompassing walnut, *Arabidopsis*, *Zea mays* and *Pilosella officinarum* hawkweed species.

1



1

Fig.3

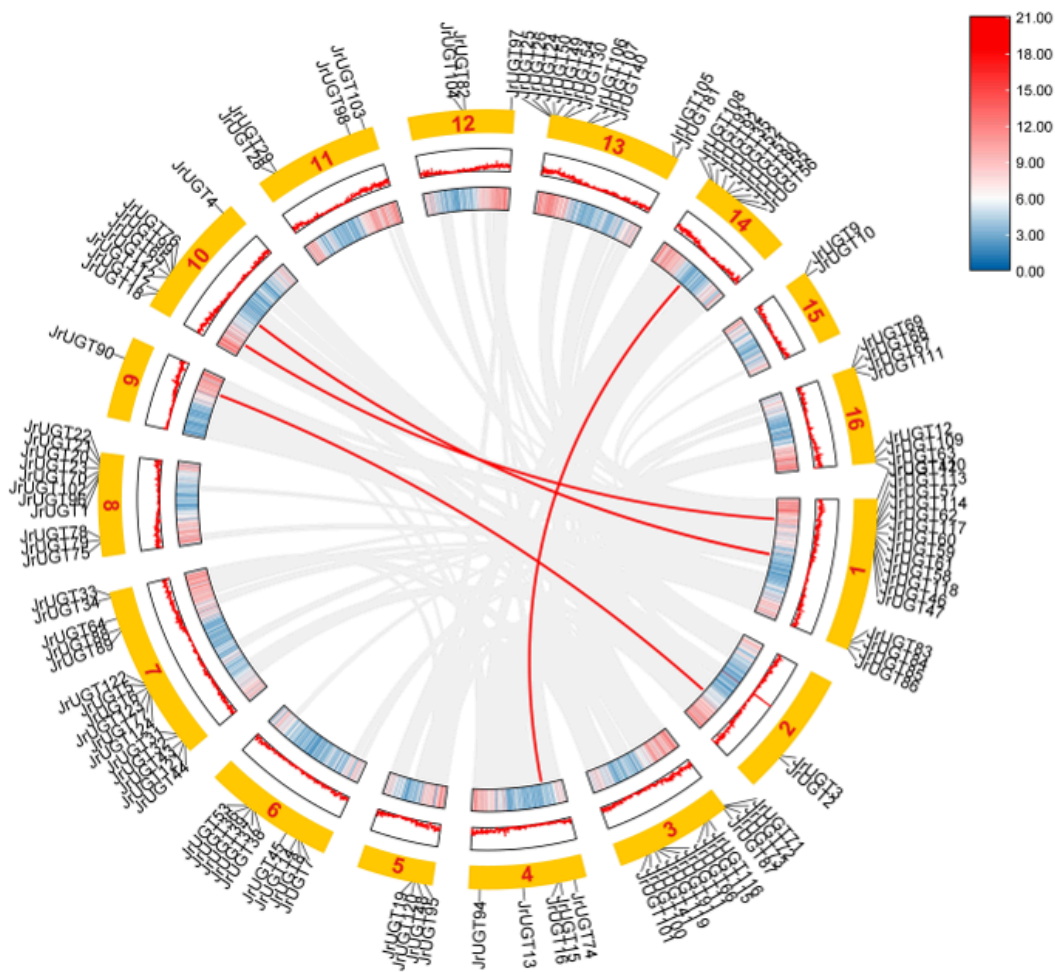


Figure 3

JrUGTs replication analysis. The yellow rectangle represents chromosomes 1-16 of the walnut, while the line map, heat map, and histogram within the rectangle depict the gene density on each chromosome. The gray lines indicate all homogenous blocks identified in the walnut genome, with red lines representing duplicate gene pairs between chromosomes.

Fig.4



Figure 4

Collinearity analysis of *UGT* genes between walnut and Arabidopsis thaliana. The gray lines depict the presence of collinear blocks within the genomes of walnut and Arabidopsis. Conversely, the red lines indicate pairs of *UGT* genes.

Fig.5

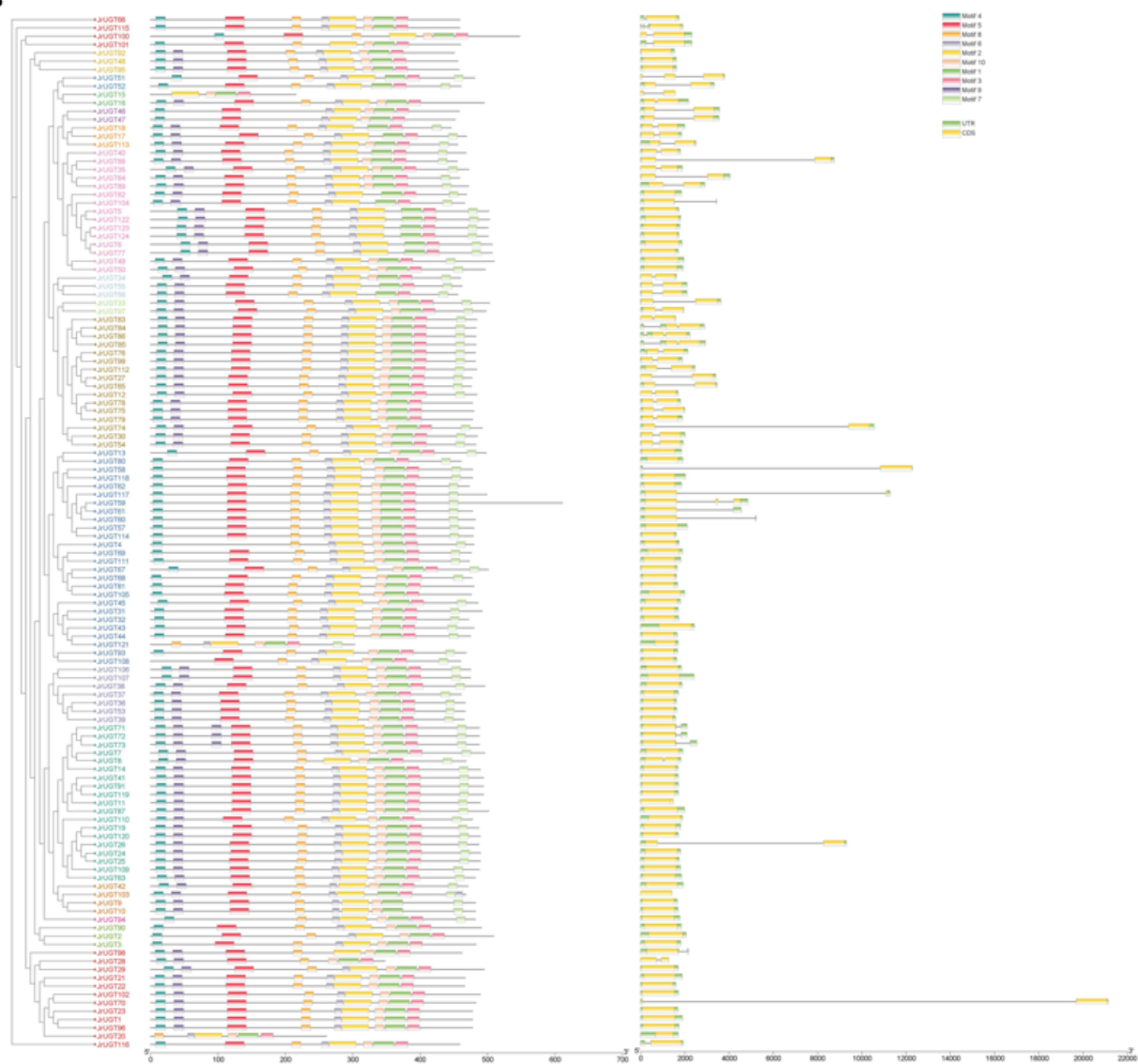


Figure 5

Conservation of sequence and gene structure analysis in the *UGT* gene family of walnut.

Fig.6

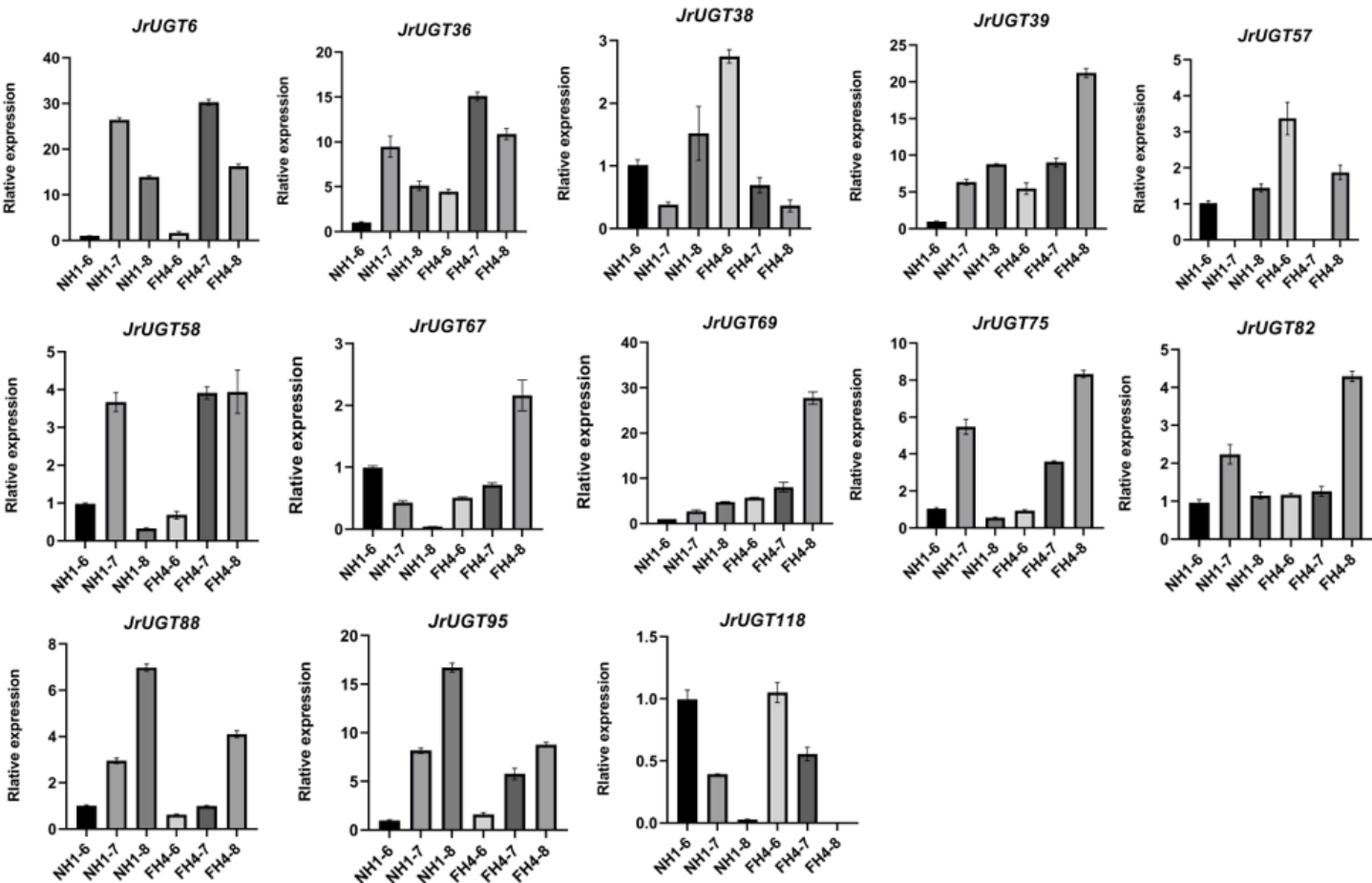


Figure 6

The expression patterns of 13 different *JrUGTs*. 6, 7 and 8 are respectively the hard core stage, the fatty stage and the mature stage of walnut.

Fig.7

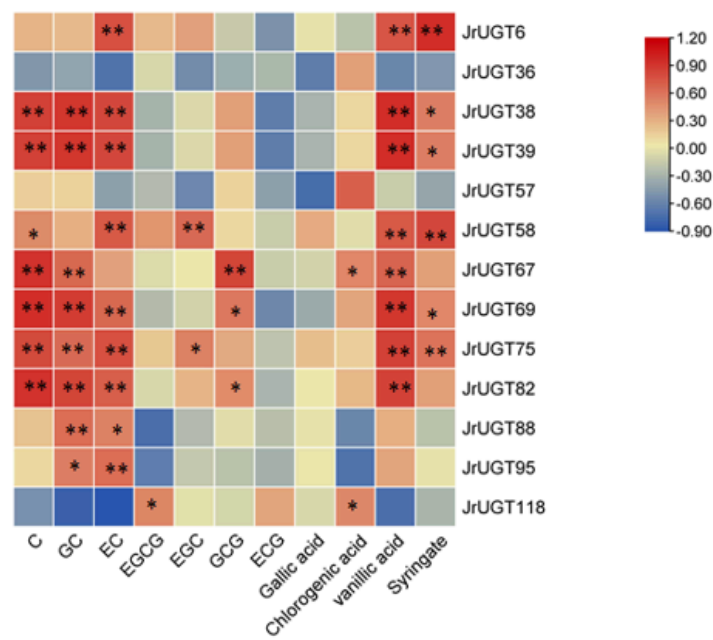


Figure 7

Analysis of the correlation between *UGT* gene expression and phenolic content in walnut. * indicates significance at the 0.05 level and ** indicates significance at the 0.01 level.

Supplementary Files

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

- [supplementaryfile.docx](#)