Genome-Wide Identification and Expression Analysis of Glycosyltransferase Gene Family 1 in *Quercus Robur* L.

Jie Zhang

North China University of Science and Technology

Li-Mei Lin

North China University of Science and Technology

Wen-Wen Cheng

North China University of Science and Technology

Xin-Song

North China University of Science and Technology

Yue-Hong Long

North China University of Science and Technology

Zhao-Bin Xing¹ (✉ xingzb@ncst.edu.cn)

North China University of Science and Technology

Research Article

Keywords: UGT, *Quercus robur* L, Glycosyltransferase, Gene family, Phylogenetic tree

Posted Date: May 6th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-474925/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Applied Genetics on July 9th, 2021. See the published version at <https://doi.org/10.1007/s13353-021-00650-3>.

Abstract

Glycosyltransferase gene family 1, also known as uridine diphosphate glycosyltransferase (UGT), is the largest glycosyltransferase family in plants, playing a vital role in their growth and development. In this study, 244 *UGT* genes with conserved PSPG motifs were identified in the genome of *Quercus robur* L. The collinearity analysis results showed that tandem repeat was the main way of *UGT* genes expansion in *Q.robur*, with 21 groups of 55 tandem repeat genes. *UGT* genes were divided into 15 subgroups A-P; group K was lost, and the gene structure and conserved domain of the same subgroup were basically the same. Cis-element analysis showed that upstream 2,000 bp promoter sequence of *UGT* genes contained light response elements, plant hormone response elements, and stress-related cis-elements, which indicated that *UGT* genes of *Q.robur* might be regulated by various metabolic pathways. In particular, some *UGTs* in group L of *Q.robur* contained a conserved promoter structure. The expression pattern analysis results demonstrated that *UGT* genes of groups B, D, E, and I were differentially expressed under *Tortrix viridana* L stress. The expression of *UGTs* in group E decreased under stress, the expression of group L increased, and that of genes in groups D and B were different. The functions of *UGT* genes in E and L groups are relatively conservative, and their functions may also conserved among species. The study results have a particular reference value for further research on the function of *Q.robur UGT* genes.

Introduce

Quercus robur L. is the deciduous tree of *Quercus* genus in Fagaceae, which is widely distributed in Europe, Asia, and North America. The wood of *Q.robur* is often used to make wine barrels (Carpena et al. 2020) and furniture (Schroeder et al. 2016). The secondary metabolites of *Q.robur* also have important medicinal value (Elansary et al. 2019); phenolic compounds have antioxidant, antibacterial, and anti-inflammatory activities (Drózdź and Pyrzynska 2018) while triterpenoids and their saponins have been widely studied for their anticancer effect (Perez et al. 2017). In the biosynthesis process of these substances, their precursors must undergo glycosyltransferase (GT) glycosylation modification to improve the physical properties and biological activities of natural product precursors and produce various functions.

Many types of GT are currently available in CAZy database (<http://www.cazy.org/GlycosylTransferases.html> April 2021), comprising 114 GT families and almost 810,000 GT enzymes, including highly homologous enzymes from plants, animals, fungi, bacteria, and viruses, which catalyze and transfer sugar from activated donor molecule to specific acceptor molecule to form glycosidic bonds. Uridine diphosphate glycosyltransferase (UGT) is the largest glycosyltransferase family in plant species (Yang et al. 2020), also known as glycosyltransferase 1 family. UGTs use uridine diphosphate (UDP) sugar donors (UDP-glucose, UDP-galactose, UDP-arabinose, UDP-rhamnose, UDP-xylose, or UDP-glucuronic acid) to catalyze the glycosylation reaction. There is a highly conserved motif PSPG at the carboxyl end of UGT protein as a unified feature which is one of the few regions with significant sequence similarity. The PSPG motif is the nucleotide-diphosphate-sugar (P05) and contains 44 amino acid residues.

The study of *UGT* genes in *Q.robur* has revealed the essential role of UDP-glycosyltransferase. Juliane Mittasch (Mittasch et al. 2014) identified the enzyme UGT84A13, which is the first step in synthesizing gallo tannins in *Q.robur* and catalyzes the production of 1-O-galloyl- β -D-glucose. Silvia Madritsch (Madritsch et al. 2019) found that UGT73C6 gene encodes flavonol-3-O-glycoside-7-O-glucosyltransferase in the research on the drought stress of *Q.robur*, attaching glucosyl residues at the 7-O-position of flavonols, kaempferol, quercetin, and their 3-O-glycoside derivatives. This gene is significantly expressed under drought conditions, which may help maintain the metabolic activity of source leaf through osmotic regulation. Studies by Birgit Kersten (Kersten et al. 2013) have shown that UGT can affect the synthesis of flavonoids, especially tannins in *Q.robur*, and plays a vital role in the resistance of *Q.robur* to diseases and insect pests. *UGT* genes could regioselectively and stereoselectively catalyze the site-specific glycosylation modification of substrates and are widely used in many fields such as medicine (Yan, Li, and Koffas 2008; Chu et al. 2017; Liang et al. 2017), industry (Wang et al. 2016) and agricultural production (Guleria and Yadav 2014). At present, there is scarce research on the *UGT* gene family of *Q.robur* at the whole genome level.

Herein, we focus on *UGT* genes in *Q.robur*, observe their position on chromosome and analyze the collinearity of genomic genes and *UGT* genes to infer their expansion in *Q.robur*. We conducted a further comparative analysis of *UGT* genes to explore their conserved domains, gene structure, and evolutionary relationships, understand the distribution of cis-elements upstream of *UGT* genes, combine transcriptome data analysis to infer the function of *UGT* genes, and provide a reference for subsequent functional identification and utilization of *UGT* genes in *Q.robur*.

Materials And Methods

Identification of UGT genes in *Q.robur*

Q.robur genome and annotation files (Plomion et al. 2018) were downloaded from OAK GENOME SEQUENCING (http://www.oakgenome.fr/?page_id=587). The hidden Markov model (HMM) profile of UGT conserved domain (PF00201) was obtained from Pfam database. The HMM file was used to query *UGT* genes in *Q.robur* using hmmsearch program of HMMER 3.0 software, with an E value threshold of 1.0 E-5. A species-specific domain model of *Q.robur* was constructed and set an E value of 1.0 E-5 to screen again. Furthermore, the online Simple Modular Architecture Research Tool (SMART) and NCBI Batch CD-Search tool were used to confirming the conserved domain for all candidate UGT protein sequences.

Chromosome location and collinearity analysis

The chromosome location was displayed using TBtools software (Chen et al. 2020). For the collinearity analysis, *Q.robur* genome files were used for sequence alignment with Basic Local Alignment Search Tool (BLAST) with a cut-off E value of 1.0 E-10. MCScanX software was employed to perform the collinearity analysis based on BLASTP results. The results were visualized using TBtools software.

Sequence alignment and phylogenetic analysis

A phylogenetic tree was constructed using UGT protein sequence of *Q.robur*. 14 UGT proteins of *Arabidopsis thaliana* L, 2 UGT proteins of *Zea mays* L, and *A.thaliana* UGT was used as an exogroup. Among them, the 2 UGTs of *Z.mays* were two UGT subgroups that did not appear in *A.thaliana*. ClustalW was used for multiple sequence alignment of UGT protein sequences with default parameters. The alignment result was used to construct an NJ phylogenetic tree by MEGA. Bootstrap resamplings (100) were used to assess the reliability of interior branches. The bootstrap value was set to 1,000. The online tool EvolView (<https://www.evolgenius.info/evolview/>) was utilized to display the phylogenetic tree.

Conserved sequence and gene structure analysis

The online MEME Suite was used to confirm the motifs of *Q.robur* UGT protein sequences with the following parameters: a maximum number of ten motifs and an optimum width of 6 to 50. The exon-intron structure analysis was performed with the software TBtools by inputting gene annotation GFF files. The sequence logo of PSPG domains was acquired using the online tool WebLogo.

Promoter region cis-acting element analysis

The 2,000 bp upstream region of initiation codon "ATG" of *UGT* genes in *Q.robur* was used to perform cis-acting element analysis with online tool PlantCARE.

(http://bioinformatics.psb.ugent.be/webtools/plantcare/html/search_CARE.html).

Transcriptome data analysis

After obtaining Silvia Madritsch's (Madritsch et al. 2019) transcriptome sequencing data of *Q.robur* under drought stress and Kersten Birgit's (Kersten et al. 2013) transcriptome sequencing data of *Q.robur* under *Tortrix viridana* L stress, differentially expressed *UGT* genes were screened. The TBtools program was used to display the heatmap of gene expression.

Results

Identification and Chromosomal Distribution of UGT genes in *Q.robur*

A total of 244 *UGT* genes were identified from the genomic sequencing data of *Q.robur*. The location of *Q.robur* *UGT* genes on chromosomes is shown in Fig. 1. A total of 219 *UGT* genes were located on 12 chromosomes, and another 25 *UGT* genes were located on unassembled chromosomal fragments. The number of *UGT* genes and tandem repeat *UGT* genes on different chromosomes of *Q.robur* is shown in Fig. 2. The most 39 *UGT* genes are located on chromosome 11, and the least *UGT* genes are located on chromosome 5, with only 6. Genes location information on chromosomes shows that *UGT* gene family in *Q.robur* is closely arranged on each chromosome, especially on chromosomes 10 and 11. Part of these tightly arranged *UGT* genes are tandem repeats, such as *P0398930.2*, *P0398960.2*, *P0398970.2*, and *P0398980.2* on chromosome 4, and other closely packed genes are not tandem repeats, such as *P0315650.2*, *P0315660.2*, and *P0315680.2* on chromosome 8.

There are 21 gene clusters on *Q.robur* chromosomes, distributed on chromosomes 1, 2, 3, 4, 5, 6, 7, 10, and 11 (Fig. 1).

Collinearity analysis showed that only 318 genes had collinearity among 28,488 genes (Fig. 3), accounting for 1.12% of all genes. The type collinear genes include genes encoding ribosomal subunit protein, genes encoding NADH plastid quinone oxidoreductase subunit, genes encoding methyltransferase, genes encoding receptor protein, and genes encoding disease resistance protein. Among them, the number of TMV resistance protein is the most, with 32. Among 244 *UGT* genes, only *P0376460.2* and *P0376630.2* on chromosome 12 were collinear.

Phylogenetic Analysis of *UGT* Sequences

A phylogenetic tree was constructed by combining 244 *Q.robur* *UGT* protein sequences, 14 *UGT* sequences identified in *A.thaliana*, and 2 *UGT* sequences identified in *Z.mays*, including 260 protein sequences. *UGTs* in *Q.robur* were divided into 15 groups (Fig. 4). *Q.robur* has O and P groups that did not exist in *A.thaliana*, and K group was lost in *Q.robur*. Two *UGTs* in *Z. mays* were clustered respectively in O and P groups; 3 *UGTs* in O group and 13 *UGTs* in P group of *Q.robur*. The number of glycosyltransferases in the 5 groups A, D, E, G, and L are larger; among them, group E has the largest number of *UGTs*, including 57 *UGTs*, which is the largest group, and group N has the least number of *UGTs*, only 1. *P0481590.2* is not clustered and is divided into a single branch.

Conserved Sequence and Gene Structure Analysis

The results of *UGT* genes structure of *Q.robur* are shown in Fig. 5b. The length of *UGT* genes is mostly less than 2,000 bp, with the longest of 5,000 bp and the shortest of 1,000 bp. There were 73 *UGT* genes with complete UTR sequence, 15 genes with 5'UTR, 22 $\geq$ 3'UTR, 134 genes without UTR, *UGTs* in groups L and G almost had no UTR sequence. *UGTs* in groups G, H, J, P, and L had long intron sequences, and all genes had no more than 3 intron fragments, while *UGTs* in groups A, B, C, D, E, O, and M had no intron sequence. The homologous *UGT* genes in each group showed a similar structure. For L group, *P0159440.2*, *P0159420.2*, *P0159330.2*, *P0159450.2*, *P0159430.2*, *P0159410.2*, *P0159310.2*, *P0159300.2*, *P0159360.2*, *P0159340.2*, *P0159320.2*, *P0296290.2*, *P0159380.2*, *P0288720.2*, *P0249750.2*, *P0760750.2*, and *P0350810.2* demonstrated great differences from other *UGTs* in the group, and they contain short introns or no intron structures.

The average length of *UGT* sequence is 433 amino acids, and N50 is 459 amino acids. Ten *UGTs* conserved motifs, named motif 1 ~ 10, including motif 1 of PSPG motif and other conserved protein motifs, were identified by online tool MEME. MEME analysis showed that unique PSPG motif (motif 1) of *UGT* gene family in *Q.robur* was located near the 3' carboxyl end. Motif 6 was lost in *UGTs* of group J, motif 7 and 10 nearly 3' carboxyl-terminal of group A *UGTs* were lost, and motif 5 close 5' terminal of some group P *UGTs* were lost (Fig. 5a). The PSPG motif of *Q.robur* (Fig. 6a) and *A.thaliana* (Fig. 6b) was constructed by intercepting all *UGTs* screened from *Q.robur* and *A.thaliana*. It was found that PSPG motif

Analysis of Cis-Acting Elements in UGT Gene Promoter Regions

The cis-acting elements of promoter sequence of *UGT* gene family of *Q.robur* can be divided into plant hormone response, involved in light responsiveness, resistant response, promoter core elements, and common cis-acting element 5 categories. Plant hormone response elements include auxin-responsive element, cis-acting element involved in gibberellin-responsiveness, cis-acting element involved in abscisic acid responsiveness, and MeJA reaction involved cis-acting regulatory elements. Cis-acting related to stress resistance include temperature, drought, hypoxia, etc.. In addition to core elements of promoter and general cis-acting elements, the number of light-responsive cis-acting elements is the largest, with 2,769. On average, there are about 10 light-responsive cis-acting elements per *UGT* gene. The resistance-related cis-acting elements are the least, only 984. The structure of promoter of *Q.robur* gene is shown in Fig. 7a. The specific *Q.robur* *UGT* genes *P0057520.2*, *P0057600.2*, *P0057620.2*, *P0057610.2*, *P0057550.2*, *P0057640.2* and *P0350810.2* have very similar promoters. The structure contains many cis-acting elements related to light response, and they belong to L group of *UGT* family of *Q.robur*. The number of different type cis-acting elements in the promoter region of *Q.robur* is displayed in Fig. 7b.

Expression analysis of UGT gene in *Q.robur*

When *Q.robur* was under *T. viridana* stress (Fig. 8a), *UGT* genes of two different *Q.robur* varieties in the same group had a similar expression pattern. For example, in the two different *Q.robur* varieties, the expression of *UGTs* in E group decreased under *T. viridana* stress, the expression of *UGTs* in L group was increased under *T. viridana* stress, and that of *UGTs* in D group was significantly different in the two *Q.robur* varieties. For *UGTs* of resistant *Q.robur* variety, *P0640850.2*, *P0703800.2*, and *P0031820.2* in D group were down-regulated. The expression of *P0170550.2*, *P0703800.2*, and *P0031820.2* of susceptible *Q.robur* variety *UGT* genes of D group was up-regulated under stress, and that of *P0640850.2* and *P0170550.2* was down-regulated.

When *Q.robur* was under drought stress, *UGTs* in group D showed different expression trends. *P0640850.2* and *P0640800.2* were down-regulated under drought stress, while *P0571810.2* was up-regulated. The genes in group D show diverse expression patterns under different stress environments (Fig. 8b).

Discussion

At present, many GT families have been identified and verified in different plants (Wu et al. 2019; Li et al. 2014). In plants, GTs are responsible for synthesizing cell wall polysaccharide chains and complex sugars (glycoproteins and glycolipids) and participating in regulating cell frameworks and physiological processes (Keegstra and Raikhel 2001). GT genes not only play an essential role in the growth and development of plants but also has paramount practical significance for humans. This study conducted

through sequence similarity, and analyzed their evolutionary amplification patterns, sequence characteristics, expression patterns, etc., which provide a basis for further research on *UGT* genes.

Expansion of the *UGT* gene family

Gene duplication is one of the main driving forces of the evolution of genomes and genetic systems, providing abundant raw materials for the evolution of genes (Moore and Purugganan 2003). Ramsey (Ramsey and Schemske 1998) once mentioned that polyploidization is the main mechanism for plants' adaptation to environment and speciation. The duplication of single genes, chromosomes, and whole genomes is the main force of plant genome structure and gene evolution (Paterson et al. 2012). Through further analysis of collinearity of *Q.robur* and chromosome location information, it was found that the level of collinearity of *Q.robur* genome was low, only 1.12%, and there was basically no collinearity among the *UGT* gene family of *Q.robur*. Simultaneously, the *UGT* gene family of *Q.robur* is tightly arranged on the chromosome, and there are more tandem repeat genes, which form numerous gene clusters on the chromosome. Therefore, it can be inferred that the main way of amplifying the *UGT* family of *Q.robur* is the tandem duplication of genes.

Evolution of *UGT* gene family

The *UGT* gene family is the largest glycosyltransferase family in plant species (Yang et al. 2020), and the largest glycosyltransferase family in *Q.robur* is also the *UGT* gene family. The *UGT* gene family in *A.thaliana* was first studied by Yi Li (Li et al. 2001), who divided *UGT* genes into 14 groups (A-N), and then found O and P groups in apple (Velasco et al. 2010), grape (Jaillon et al. 2007), soybean (Schmutz et al. 2010), rice (Tanaka et al. 2008) and corn (Li et al. 2014). A total of 244 *UGTs* identified in *Q.robur* were divided into 15 groups, including O and P groups. Group K was lost in *Q.robur* (Fig. 4), and group J, which was closer to group K in *A.thaliana*, had 10 *UGTs*. There are both 2 *UGT* genes in K and J in *A.thaliana* (Li et al. 2001), in *Z.mays* (Li et al. 2014), there are 2 and 4 *UGTs* in K and J group, and in *Malus domestica* (Zhou et al. 2017), there are 8 and 14 *UGTs* in group K and group L, respectively. It can be found that J group in the plant *UGT* family is larger than K group. It can be speculated that *UGT* genes of K group may have evolved from J group; in *Q.robur*, K group has not yet been completely separated from J group. *Q.robur* *UGT* gene family A, D, G, E, and L expand faster. Caputi Lorenzo (Caputi et al. 2012) proposed that E group genes in different plant species account for 20–25% of *UGT* family. *UGT* expansion in E group of *Q.robur* was also the fastest, accounting for 23% of *UGT*, consistent with Caputi Lorenzo's research.

Structure and protein conserved motifs of *UGT* gene

The presence of UTR can regulate the binding efficiency of RNA polymerase during translation and promote gene expression through translation initiation factors (Kim et al. 2020). In *UGT* gene family, 110 genes have UTR, accounting for 45% of all *UGTs*. The *UGT* genes of L and G groups lack UTR structure (Fig. 5), but for L group, many cis-elements are related to light response in the promoter part (Fig. 7), indicating a significant difference in the expression pattern of *UGT* gene family. This difference is closely

The sequence of UGT protein in 15 groups showed obvious conservation and inter-group specificity. Conserved structural sequence is the prerequisite for conservation of biological function. The PSPG motif of *Q.robur* UGT is highly similar to that of *A.thaliana* UGT (Fig. 6). There is a conservative "WAPQVEVLSPAVGGFVTHCGWNSTLESLVAGVPMICWPLFADQ" motif at the carboxyl end of the protein sequence of the 244 *UGT* genes screened, consistent with previous research (Paquette, Møller, and Bak 2003). The conservative motifs of different *UGT* genes of *Q.robur* also show a high degree of similarity (Fig. 5), closely related to the general function of glycosyltransferase that can transport active sugar molecules to specific receptors.

Cis-Acting Elements of UGTs in *Q.robur*

The conserved protein structure produces conserved functions, and this conservation is also reflected in the promoter structure. Cis-acting elements can combine with transcription factors to regulate gene expression. The similar expression pattern caused by this similar promoter structure predicts the conservation of gene function to a certain extent. The *UGT* genes *P0057520.2*, *P0057600.2*, *P0057620.2*, *P0057610.2*, *P0057550.2*, *P0057640.2* and *P0350810.2*, which belong to L group of *Q.robur*, have very similar light response-related promoter structures (Fig. 6a). In *A.thaliana*, L group of *UGT* gene family includes UGT84A1-4, UGT84B1-2, UGT75B1-2, UGT75D1, and UGT75C1. UGT84A1-A4 was related to the metabolism of hydroxycinnamates (Meissner et al. 2008), and UGT75C1 has been identified as an anthocyanin 5-O-glucosyltransferase (Tohge et al. 2005). Studies have shown that anthocyanin and other secondary metabolites in plant cells can protect plants from UV damage (Jaakola 2013). Dirk Meißner (Meissner et al. 2008) found that when *A.thaliana* was exposed to UV-B stress, the content of UGT84A1-A4 enzyme-mediated UV-shielding compound sinapate (hydroxycinnamates) increased instantaneously. It accumulates in the bubble and can play a protective role under short-term UV-B radiation conditions. It can be speculated that expression of some *UGTs* in L group in *Q.robur* might be regulated by light to enhance adaptability under light stress.

UGT gene expression analysis

The *UGT* genes in E group in *A.thaliana* are closely related to cell wall development and lignification, such as UGT72B1 (Lin et al. 2016) and UGT72E1-3 (Baldacci-Cresp et al. 2020). UGT84B1, UGT84B2, UGT75B1, and UGT75B2 in *A.thaliana* can catalyze the production of glycosylation products from plant hormone indole-3-acetic acid (IAA), which in turn regulates plant growth and development (Jackson et al. 2001). If functions of *UGT* genes in L and E groups of *Q.robur* are similar to those in *A.thaliana*, up-regulation, and down-regulation of *UGT* genes of L and E groups in *Q.robur* have a certain correlation under the stress of *T. viridana*. Under this stress, *UGT* genes expression in E group was up-regulated, which causes the content of auxin IAA to decrease in *Q.robur* tissue due to glycosylation. The down-regulation of gene expression in L group directly caused cell wall formation and development to be blocked. The growth of this tissue is stagnated under stress. The substance and energy are used for growing other parts of *Q.robur*. E and L groups are the two largest subfamilies in the *UGT* gene family of *Q.robur* and their importance for plant growth and development may be an essential reason for the

conservation of functions of E and L subfamilies of *UGT* gene family among species. *UGT* genes in group D show different expression patterns under different stress (Fig. 8), and the functions of group D *UGT* genes may be quite differentiated.

Declarations

Funding The project was supported by grants from Natural Science Foundation of Hebei Province (H2020209033) and, Hebei Education Department sponsored scientific research projects (ZD2019075).

Conflict of interest The authors declare that they have no competing interests.

Authors' contributions The first draft of the manuscript was written by Jie Zhang and all authors commented on previous versions of the manuscript. Material preparation, data collection and analysis were performed by Li-Mei Lin and Jie zhang. Zhao-bin Xing contributed to the study conception and design. All authors read and approved the final manuscript.

References

1. Baldacci-Cresp F, Le Roy J, Huss B et al (2020) UDP-glycosyltransferase 72E3 plays a role in lignification of secondary cell walls in Arabidopsis. *Int J Mol Sci* 21:6094. <https://doi.org/10.3390/ijms21176094>
2. Bowles D, Isayenkova J, Lim EK et al (2005) Glycosyltransferases: managers of small molecules. *Curr Opin Plant Biol* 8:254–263. <https://doi.org/10.1016/j.pbi.2005.03.007>
3. Caputi L, Malnoy M, Goremykin V et al (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. *Plant J* 69:1030–1042. <https://doi.org/10.1111/j.1365-313X.2011.04853.x>
4. Carpena M, Pereira AG, Prieto MA et al (2020) Wine aging technology: fundamental role of wood barrels. *Foods* 9:1160. <https://doi.org/10.3390/foods9091160>
5. Chen CJ, Chen H, Zhang Y et al (2020) TBtools: An integrative toolkit developed for interactive analyses of big biological data, *Molecular Plant*, 13: 1194 – 202
6. Chu LL, Pandey RP, Lim HN et al (2017) Synthesis of umbelliferone derivatives in *Escherichia coli* and their biological activities. *J Biol Eng* 11:15. <https://doi.org/10.1186/s13036-017-0056-5>
7. Drózdź P, Krystyna et al (2018) Assessment of polyphenol content and antioxidant activity of oak bark extracts. *Eur J Wood Wood Product* 76:793–795. <https://doi.org/10.1007/s00107-017-1280-x>
8. Elansary HO, Szopa A, Kubica P et al (2019) Polyphenol profile and pharmaceutical potential of *Quercus* spp. *Bark Extracts Plants (Basel)* 8:486. <https://doi.org/10.3390/plants8110486>
9. Guleria P, Yadav SK (2014) Overexpression of a glycosyltransferase gene SrUGT74G1 from *Stevia* improved growth and yield of transgenic Arabidopsis by catechin accumulation. *Mol Biol Rep*

- 41:1741–1752. <https://doi.org/10.1007/s11033-014-3023-y>
10. Jaakola L (2013) New insights into the regulation of anthocyanin biosynthesis in fruits. *Trends Plant Sci* 18:477–483. <https://doi.org/10.1016/j.tplants.2013.06.003>
 11. Jackson RG, Lim EK, Li Y et al (2001) Identification and biochemical characterization of an *Arabidopsis* indole-3-acetic acid glucosyltransferase. *J Biol Chem* 276:4350–4356. <https://doi.org/10.1074/jbc.M006185200>
 12. Jaillon O, Aury JM, Noel B et al (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449:463–467. <https://doi.org/10.1038/nature06148>
 13. Keegstra K, Raikhel N (2001) Plant glycosyltransferases, *Curr Opin Plant Biol*, 4: 219–224. [https://doi.org/10.1016/s1369-5266\(00\)00164-3](https://doi.org/10.1016/s1369-5266(00)00164-3)
 14. Kersten B, Ghirardo A, Schnitzler JP et al (2013) Integrated transcriptomics and metabolomics decipher differences in the resistance of pedunculate oak to the herbivore *Tortrix viridana* L. *BMC Genom* 14:737. <https://doi.org/10.1186/1471-2164-14-737>
 15. Kim B, Kim HM, Kang MK et al (2020) 5'-UTR and ORF elements, as well as the 3'-UTR regulate the translation of Cyclin. *Biochem Biophys Res Commun* 527:968–973. <https://doi.org/10.1016/j.bbrc.2020.04.151>
 16. Li Y, Baldauf S, Lim EK et al (2001) Phylogenetic analysis of the UDP-glycosyltransferase multigene family of *Arabidopsis thaliana*. *J Biol Chem* 276:4338–4343. <https://doi.org/10.1074/jbc.M007447200>
 17. Li Y, Li P, Wang Y et al (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. Liang H, Hu Z, Zhang T et al (2017) Production of a bioactive unnatural ginsenoside by metabolically engineered yeasts based on a new UDP-glycosyltransferase from *Bacillus subtilis*. *Metab Eng* 44:60–69. <https://doi.org/10.1016/j.ymben.2017.07.008>
 19. Lin JS, Huang XX, Li Q et al (2016) UDP-glycosyltransferase 72B1 catalyzes the glucose conjugation of monolignols and is essential for the normal cell wall lignification in *Arabidopsis thaliana*. *Plant J* 88:26–42. <https://doi.org/10.1111/tpj.13229>
 20. Madritsch S, Wischnitzki E, Kotrade P et al (2019) Elucidating drought stress tolerance in european oaks through cross-species transcriptomics, *G3 (Bethesda)*, 9: 3181–3199. <https://doi.org/10.1534/g3.119.400456>
 21. Meissner D, Albert A, Böttcher C et al (2008) The role of UDP-glucose:hydroxycinnamate glucosyltransferases in phenylpropanoid metabolism and the response to UV-B radiation in *Arabidopsis thaliana*. *Planta* 228:663–674. <https://doi.org/10.1007/s00425-008-0768-3>
 22. Mittasch J, Böttcher C, Frolova N et al (2014) Identification of UGT84A13 as a candidate enzyme for the first committed step of gallotannin biosynthesis in pedunculate oak (*Quercus robur*), <https://doi.org/10.1016/j.phytochem.2013.11.023>

23. Moore RC, Purugganan MD (2003) The early stages of duplicate gene evolution. *Proc Natl Acad Sci U S A* 100:15682–15687. <https://doi.org/10.1073/pnas.2535513100>
24. Paquette S, Møller BL, Bak S (2003) On the origin of family 1 plant glycosyltransferases. *Phytochemistry* 62:399–413. [https://doi.org/10.1016/s0031-9422\(02\)00558-7](https://doi.org/10.1016/s0031-9422(02)00558-7)
25. Paterson AH, Wendel JF, Gundlach H et al (2012) Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* 492:423–427. <https://doi.org/10.1038/nature11798>
26. Perez AJ, Pecio L, Kowalczyk M et al (2017) Triterpenoid components from oak heartwood (*Quercus robur*) and their potential health benefits. *J Agric Food Chem* 65:4611–4623. <https://doi.org/10.1021/acs.jafc.7b01396>
27. Plomion C, Aury JM, Amselem J et al (2018) Oak genome reveals facets of long lifespan. *Nat Plants* 4:440–452. <https://doi.org/10.1038/s41477-018-0172-3>
28. Ramsey J, Schemske DW (1998) Pathways, mechanisms, and rates of polyploid formation in flowering plants. *Annu Rev Ecol Syst* 29:467–501. <https://doi.org/10.1146/annurev.ecolsys.29.1.467>
29. Schmutz J, Cannon SB, Schlueter J et al (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463:178–183. <https://doi.org/10.1038/nature08670>
30. Schroeder H, Cronn R, Yanbaev Y et al (2016) Development of molecular markers for determining continental origin of wood from white oaks (*Quercus L. sect. Quercus*). *PLoS One* 11:e0158221. <https://doi.org/10.1371/journal.pone.0158221>
31. Tanaka T, Antonio BA, Kikuchi S et al (2008) The Rice annotation project database (RAP-DB): 2008 update, *Nucleic Acids Res*, 36: D1028-D1033. <https://doi.org/10.1093/nar/gkm978>
32. Tohge T, Nishiyama Y, Hirai MY et al (2005) Functional genomics by integrated analysis of metabolome and transcriptome of *Arabidopsis* plants over-expressing an MYB transcription factor. *Plant J* 42:218–235. <https://doi.org/10.1111/j.1365-313X.2005.02371.x>
33. Velasco R, Zharkikh A, Affourtit J et al (2010) The genome of the domesticated apple (*Malus × domestica* Borkh.). *Nat Genet* 42:833–839. <https://doi.org/10.1038/ng.654>
34. Wang Y, Chen L, Li Y et al (2016) Efficient enzymatic production of rebaudioside A from stevioside. *Biosci Biotechnol Biochem* 80:67–73. <https://doi.org/10.1080/09168451.2015.1072457>
35. Wu A, Hao P, Wei H et al (2019) Genome-wide Identification and characterization of glycosyltransferase family 47 in Cotton. *Front Genet* 10:824. <https://doi.org/10.3389/fgene.2019.00824>
36. Yan Y, Li Z, Koffas MA (2008) High-yield anthocyanin biosynthesis in engineered *Escherichia coli*. *Biotechnol Bioeng* 100:126–140. <https://doi.org/10.1002/bit.21721>
37. Yang C, Li C, Wei W et al (2020) The unprecedented diversity of UGT94-family UDP-glycosyltransferases in *Panax* plants and their contribution to ginsenoside biosynthesis. *Sci Rep* 10:15394. <https://doi.org/10.1038/s41598-020-72278-y>

38. Zhou K, Hu L, Li P et al (2017) Genome-wide identification of glycosyltransferases converting phloretin to phloridzin in *Malus* species. *Plant Sci* 265:131–145.
<https://doi.org/10.1016/j.plantsci.2017.10.003>

Figures

Figure 1

The location information of *Q.robur* UGT genes on chromosome. Tandem duplicated UGT genes are

Figure 2

Distribution of UGT genes and tandem repeat UGT genes on *Q.robur* chromosomes.

Figure 3

Collinearity of *Q.robur* genome. The red lines indicate UGT genes, and the green lines indicate other genes.

Figure 4

UGT proteins phylogenetic tree of *Q.robur*. The 260 protein sequences, including 244 UGT sequences of *Q.robur*, 14 UGT sequences of *A. thaliana*, and 2 UGT sequences of *Z.mays*, were used to construct a phylogenetic tree with a bootstrap value of 1,000 by MEGA.

Figure 5

The evolutionary relationship, motif structure, and gene structure of *Q.robur* UGTs. a. The UGT protein sequences. b. The UGT gene structure.

Figure 6

PSPG motif drawn by 244 UGT genes in Q.robur and PSPG motif drawn by 116 UGT genes in A.thaliana.
a. The Q. robur PSPG motif. b. The A.thaliana PSPG motif.

Figure 7

Promoter structure of *Q.robur* UGT genes. a. The position of various cis-acting elements in the promoter region of 2,000 bp upstream of UGT genes. The black outlined box is UGT genes with similar promoter structures. b. The number of different types of cis-acting elements in the promoter region.

Figure 8

The expression of UGT genes in leaves of *Q.robur* under *T. viridana* stress and drought stress. a. The expression of UGT genes in leaves of *Q.robur* under *T. viridana* stress. T represents the *Q.robur* which is not easy to be infected by *T. viridana*; S represents the *Q.robur* is easy to be infected by *T. viridana*; co means control group, fed means the samples of *Q.robur* treated with *T. viridana*; b. The expression of UGT genes in leaves of *Q.robur* under drought stress; CO-1, 2 means *Q.robur* control group, DS-1, DS-2, DS-3 indicate *Q.robur* is under drought stress.