Research Square

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

Analysis of UGT Gene Family and Functional Involvement in Drought and Salt Stress Tolerance in Phoebe bournei

Zhening Zhu
Fujian Agriculture and Forestry University
Yanzi Zhang
Fujian Agriculture and Forestry University
Jingshu Li
Fujian Agriculture and Forestry University
Sijia Yang
Fujian Agriculture and Forestry University
Jinxi Pan
Fujian Agriculture and Forestry University
Hengfeng Guan
Fujian Agriculture and Forestry University
Almas Bakari
Fujian Agriculture and Forestry University
Shipin Chen
Fujian Agriculture and Forestry University
Shijiang Cao (000q813040@fafu.edu.cn)
Fujian Agriculture and Forestry University

Research Article

Keywords: Phoebe bournei, UDP-glycosyltransferase, expression analysis, bioinformatics analysis

Posted Date: July 6th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3093086/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

Background

Uridine diphosphate glycosyltransferases (UDP-GTs, UGTs), regulated by *UGT* genes, play a pivotal role in glycosylation, a prevalent compound modification pathway observed in plant secondary metabolism. Glycosylation is known to contribute significantly to hormone homeostasis, plant detoxification processes, and secondary metabolite production in plants. Despite extensive studies on the *UGT* gene family in numerous species, a comprehensive investigation of the *UGT* family in *Phoebe bournei* remains unreported. Thus, we conducted a systematic analysis to identify 156 *PbUGT* genes and provide a comprehensive examination of this family by investigating homology, conserved protein regions, gene structure, chromosomal location, and tissue-specific expression patterns.

Results

We identified 156 *UGT* genes within the entire *P. bournei* genome, all of which contained the PSPG box. The *PbUGT* family consists of 14 subfamilies, consistent with *Arabidopsis thaliana*. We observed varying expression levels of *PbUGT* genes across different tissues in *P. bournei*, with the following average expression hierarchy: leaf > stem xylem > stem bark > root xylem > root bark. Covariance analysis revealed stronger covariance between *P. bournei* and closely related species. Additionally, PbUGT genes exhibited differential expression under drought and salt stresses, with specific expression patterns observed under each stress condition.

Conclusions

Our findings shed light on the transcriptional response of *PbUGT* factors to drought and salt stresses, thereby establishing a foundation for future investigations into the role of *PbUGT* transcription factors.

Background

Glycosylation, a widespread compound modification pathway in plant secondary metabolism, is a crucial transformation process in organisms. It serves as the final stage for many biosynthetic products, altering the activity, stability, and solubility of receptors [1]Glycosylation plays a vital role in plant hormone balance, detoxification processes, and secondary metabolite production. Glycosyltransferases (GTs) are responsible for plant glycosylation and constitute a highly diverse multi-gene family present in various organisms. These enzymes catalyze the specific transfer of glycosyl groups from donors to receptors, forming α- and β-glycosidic bonds and generating oligosaccharides, polysaccharides, complex sugars (such as glycoproteins and glycolipids), and glycoside compounds [2]. The most prevalent glycosyl donor in plants is UDP-glucose, while receptors include glycolipids, proteins, and nucleic acids.

The carbohydrate active enzyme database (CAZy) classifies glycosyltransferases (GTs) from different species into 114 gene families based on amino acid sequence similarity, substrate specificity, catalytic mechanism, signal membrane, and conformation of the glycosidic chain. The GT1 branch represents the largest division, indicating its close functional association with plants. Within the GT1 branch, there exists a subclass known as UDP-glucosyltransferases (UGTs) due to their utilization of UDPG as the glycosyl donor [3]. One prominent characteristic

of the plant UGT family is the presence of a highly conserved segment, known as the PSPG box, comprising 44 amino acids located at the protein's C-terminus [4]. This conserved motif facilitates the recognition and interaction with uridine diphosphate glycosyl donors. Conversely, the N-terminal sequence, which is considerably variable and less conserved, governs the recognition and binding of various receptors, contributing to the vast substrate diversity and complex structures of plant *UGT* genes, such as their involvement in the glycosylation of diverse compounds including hormones, terpenoids, and sterols [5].

The *UGT* gene family has been extensively identified and classified in various plants, including *Arabidopsis thaliana* [6], *Linum usitatissimum* [7], *Triticum aestivum* [8], *Zea mays* [9], *Gossypium hirsutum* [10], and *Prunus persica* [11]. *Arabidopsis thaliana*, a model plant, possesses 122 *UGT* genes, among which eight are pseudogenes. These genes are categorized into 14 distinct groups (A-N) based on their amino acid sequences [6]. Subsequent studies have revealed the presence of O and P groups in higher plants [12]. *Oryza sativa* has a total of 609 *UGT* genes, divided into 40 groups along with an unknown class [13]. *Citrus grandis* exhibits 145 *UGT* genes distributed among 16 groups [14], while *Zea mays* contains 147 *UGT* genes classified into 17 groups, with 18 genes shared with *Arabidopsis thaliana* and two genes shared with *Oryza sativa* [9]. The functional roles of the *UGT* gene family have been validated in various plants, including hormone balance regulation, localization control, secondary metabolite synthesis, and detoxification and defense responses [15].

Phoebe bournei, a species of the Lauraceae family, is characterized by a straight trunk and limited branches. It serves as an excellent timber species and is endemic to China [16]. Noteworthy for its unique fragrance, *P. bournei* holds significant value as a timber tree. Its wood exhibits strong resistance to corrosion, making it resistant to cracking and bending during the drying process. Consequently, it possesses high economic worth and finds applications in shipbuilding and furniture production [17]. However, due to excessive logging driven by its economic value and the adverse effects of environmental factors such as soil fertility degradation, *P. bournei* has experienced a decline in its wild population and is now listed as a nationally secondary protected plant.

While several gene families in other plants have been extensively studied, limited research has been conducted on the gene family of *P. bournei*. This dearth of studies can be attributed to the relatively recent sequencing of the *P. bournei* genome. The complete genome data of *P. bournei* was initially obtained by Liu Zhongjian and Chen Shiping from Fujian Agriculture and Forestry University in 2020 [18]. However, this study did not unveil the chromosomal-level information. Subsequently, in 2022, the genome of *P. bournei* was deciphered by the Tong Zaikang team from Zhejiang Agriculture and Forestry University, shedding light on the synthesis mechanism of *P. bournei* sesquiterpenes [19].

The gene families in *P. bournei* primarily include commonly found families like *WRKY* [20], *NF-Y* [21], and *PLR* [22], while the *UGT* gene family has received limited attention. Previous studies have primarily focused on whole gene family analyses, with few investigations on stress responses. Furthermore, existing research primarily addresses nutrient deficiency stress, neglecting drought and salt stress. Therefore, our study aims to elucidate the response mechanisms of *PbUGT* genes under drought and salt stress conditions.

Results

Identification and physicochemical analysis of PbUGT gene members

Through rigorous verification using various methods, a comprehensive set of 156 *PbUGT* genes was successfully identified within the entirety of the genome. These *UGT* genes were subsequently designated *PbUGT1-156* (Table 1)

based on their chromosomal distribution. To gain further insights into the physicochemical properties of the UGT proteins, ProtParam, MEME, and other online tools were employed for analysis. The results revealed that the amino acid count of the 156 PbUGT proteins ranged from 132 aa (PbUGT66) to 1019 aa (PbUGT89), with an average length of 458 aa. The molecular weight of the proteins varied between 14948.04 Da (PbUGT66) and 113398.82 Da (PbUGT89), with an average weight of 50994.75 Da. The isoelectric points spanned a range from 4.62 to 9.8, with an average value of 5.87. Specifically, there were 136 acidic proteins (PI < 6.5), 12 neutral proteins (6.5 < PI < 7.5), and 8 basic proteins (PI > 7.5). Subcellular localization analysis of PbUGT proteins indicated that 88 proteins were localized within the chloroplast, 43 in the cytoplasm, and 15 in the nucleus, while the remaining proteins were distributed across various organelles such as vacuoles, peroxisomes, and mitochondria. Furthermore, instability coefficient analysis unveiled that 46 proteins were classified as stable, whereas the remaining 110 proteins exhibited instability. Notably, the maximum instability coefficient reached a significant value of 73.59, suggesting that the majority of PbUGT proteins were characterized as unstable acidic proteins.

Table 1
Physical and chemical properties of PbUGT family members

Name	Average molecular weight	Theoretical pI	Asp + Glu	Arg + Lys	Instability index	Aliphatic index	GRAVY
Distribution range	14948.04-113398.82	4.62–9.8	17–127	11–110	32.1-72.59	71.5-102.3	-0.467-0.273
Average value	50994.75	5.87	54	44	44.31	88.18	-0.121
Minimum value member	PbUGT66	PbUGT140	PbUGT154	PbUGT154	PbUGT144	PbUGT136	PbUGT66
Maximum value member	PbUGT89	PbUGT1	PbUGT89	PbUGT9	PbUGT122	PbUGT75	PbUGT154

Chromosome Localization Analysis of the PbUGT Gene Family

To determine the precise chromosomal positions of the *UGT* genes, we utilized MapChart software to construct a visual representation of their distribution. Our analysis revealed that the 156 *UGT* genes were dispersed across all 12 chromosomes. Notably, chromosome 10 harbored the highest number of *PbUGT*s, with a total of 32 genes, followed by chromosome 2 with 28 *PbUGT*s. Furthermore, the distribution pattern of *UGT* gene members within different subfamilies of *P. bournei* displayed irregularities, characterized by the presence of gene clusters. These clusters varied in size, ranging from two to 13 genes, with the largest cluster residing on chromosome 10. We further examined the chromosomal distribution within different gene groups. Chromosome 2 contained 28 genes belonging to eight distinct groups. Among these, group A exhibited the greatest number of members in the *P. bournei UGT* family, totaling 36 genes. Specifically, 13 members from this group formed a cluster on chromosome 7, while four members clustered on chromosomes 2 and 4. The remaining six members were dispersed across chromosomes 1, 2, 5, and other regions encompassing chromosomes 8 to 12 (Fig. 1). The uneven distribution of *PbUGT* genes across chromosomes indicates the presence of genetic variations during the evolution of *P. bournei*.

Gene Structure, Protein Conserved Motif, and Exon-Intron Analysis of the PbUGT Gene Family

To gain further insights into the conserved domain characteristics of the *UGT* family in *Phoebe bournei*, we employed the online tool MEME to identify and assign 10 distinct motifs, designated as motifs 1 to 10 (Fig. 2). The motif analysis revealed that the number of motifs present in each *PbUGT* family member ranged from three to 13. Notably, 114 out of the 156 members exhibited a combination of motifs 4 to 9. This unique motif combination serves as a distinctive feature of the family, setting it apart from other gene families and implying its close association with specific functional and performance sequences. Additionally, a notable observation was that members belonging to the same evolutionary group exhibited a high degree of similarity in the distribution of conserved motifs. This suggests that members classified within the same group are likely to share similar functions due to the similarity in motif distribution. Domain analysis unveiled the presence of four conserved domains, namely Glycosyltransferase, GTB-type superfamily, GT1-Gtf-like, and PLN02448 domains, in all 156 *UGT* members. This observation suggests that the composition of these three conserved motifs is intricately linked to the synthesis and regulation of glycosyltransferases.

The *PbUGT* gene structure and exon-intron organization were further analyzed and visualized (Fig. 2). Among the 156 members, a total of 102 (65.4%) members were found to possess introns, distributed across all 14 groups. Group G exhibited the highest number of introns, with 29 members, followed by group A with 17 members. Notably, introns were present in all groups except for groups F, I, J, K, M, and N. The distribution of introns can provide valuable insights into the functional properties of each gene. Therefore, studying the differences in intron positions can offer further exploration into the distinct biological characteristics of each member within the gene family.

Phylogenetic Analysis of the PbUGT Gene Family

Previous studies have classified *UGT* genes in *Arabidopsis* into 14 distinct groups, labeled A to N, with the identification of three additional groups, O, P, and Q, in higher plants. To elucidate the functional attributes of each *PbUGT* family gene and determine its phylogenetic group, we obtained 122 sequences from *A. thaliana* via the *Arabidopsis* resource website. Subsequently, we constructed a phylogenetic tree by incorporating these sequences with the 156 identified *P. bournei* sequences. The phylogenetic tree revealed the presence of 14 *A. thaliana* groups within *P. bournei*, with an uneven distribution across these groups (Fig. 3). Notably, the majority of *PbUGT* genes were found in groups A, G, E, and L, with the largest representation in group A, consisting of 36 members. Conversely, groups K, M, and N each contained only one member. The differences observed among various subfamilies may be linked to the evolutionary trajectory and functional diversity within *P. bournei.*

Tissue Expression Profile Analysis of the PbUGT Gene Family

By leveraging the original transcriptome sequencing data obtained from five distinct tissues in *P. bournei* (PRJNA628065), we constructed an expression heatmap to examine the tissue-specific expression patterns of the 156 *PbUGT* genes (Fig. 4). The heatmap displayed expression levels ranging from low (white) to high (dark red), highlighting considerable variations in the expression patterns across different tissues. Among the genes analyzed, *PbUGT1*, *PbUGT21*, *PbUGT26*, *PbUGT27*, *PbUGT35*, *PbUGT55*, *PbUGT84*, *PbUGT104*, *PbUGT117*, *PbUGT121*, *PbUGT145*, and *PbUGT154* lacked corresponding expression data in the reported transcriptome, suggesting their limited involvement in the growth of *P. bournei*. In contrast, *PbUGT101* and *PbUGT136* exhibited significantly higher expression levels (FPKM > 15) in all tissues, indicating their crucial roles in the growth and development of *P. bournei*. On average, the expression levels ranked as follows: leaf > stem xylem > stem epidermis > root xylem > root epidermis. Among the fourteen subfamilies, the B subfamily displayed the highest average expression level, surpassing other subfamilies by a substantial margin. Conversely, the C, D, and K subfamilies exhibited notably low

overall expression levels (FPKM < 1). The differential expression of *PbUGT* genes across various tissues signifies functional disparities among genes within distinct subfamilies.

Synteny Analysis of Homologous UGT Genes Among Phoebe bournei, Arabidopsis thaliana, Populus trichocarpa, and Cinnamomum camphora

To investigate the collinearity relationships within the *UGT* gene family among *Phoebe bournei*, *Arabidopsis thaliana*, *Populus trichocarpa*, and *Cinnamomum camphora*, a comprehensive genome-wide collinearity analysis was conducted. The aim was to unveil the interconnections within the *UGT* family. As depicted in Fig. 5, a total of 13 *PbUGT* genes on eight chromosomes displayed collinearity with *UGT* genes from *A. thaliana*, while 39 *UGT* genes on 10 chromosomes exhibited collinearity with *UGT* genes from *P. trichocarpa*. Additionally, 60 *UGT* genes on all 12 chromosomes displayed collinearity with *UGT* genes from *Cinnamomum camphora*. The collinearity analysis revealed that species with closer genetic relationships exhibited stronger collinearity within the same gene family. This observation suggests a plausible association with species evolution, where these genes might have arisen earlier, with each gene family demonstrating homology prior to species differentiation. Over time, as species diverged, these homologous gene relationships gradually diminished.

Duplication Analysis of PbUGT Genes in Chromosomal Replication

To elucidate the expansion and evolutionary mechanisms underlying the *PbUGT* gene family, potential gene duplication events in the *P. bournei* genome were investigated. The entire *P. bournei* genome was compared using MCScanX software, leveraging amino acid sequence homology. The analysis revealed the presence of 14 tandem repeat gene clusters within the *UGT* gene family members (Fig. 6), indicating that tandem repeats played a pivotal role in the gene family's expansion. Additionally, the study examined the ratio of nonsynonymous substitutions (Ka) to synonymous substitutions (Ks) between the duplicated genes. The results demonstrated that the ratio was significantly below 1, implying a negative selection effect on the *PbUGT* gene during the evolutionary process, indicative of purifying selection.

Enrichment Analysis of Stress, Hormone, and Light-Responsive Cis-Elements in the Promoter Region of Phoebe bournei UGT Genes

Cis-acting elements serve as indicators of gene function and predictors of transcriptional regulation patterns. The promoter regions of *PbUGT* genes encompass numerous cis-acting elements with distinct functions, which can be categorized into stress response, hormone response, and light response. In total, seven, 35, and 11 cis-elements associated with stress response, hormone response, and light response were identified across all *PbUGT* promoter sequences, respectively. Among them, the G-box (a light response cis-element) exhibited the highest frequency, occurring 366 times in the promoter sequences of 131 genes. It was followed by ABRE (a hormone response cis-element), which appeared 326 times in the promoter sequences of 124 genes. Previous studies have indicated that cis-element responses in various plants are involved in diverse stresses and hormone regulation. Hence, *PbUGT* genes may play a role in stress, hormone, and light regulation during the growth and development of *P. bournei*. This finding provides valuable insights for subsequent stress treatments in this experimental analysis.

Screening and Verification of Stress Resistance Genes in PbUGT Genes

Based on the cis-acting elements, five genes-*PbUGT14*, *PbUGT15*, *PbUGT56*, *PbUGT60*, and *PbUGT154*—displaying the highest number of elements corresponding to adversity were identified. Subsequently, their transient expression levels were verified in response to various adversity stresses using RT-qPCR. The expression level analysis of *PbUGT*

genes under 1.9 M NaCl salt stress revealed that *PbUGT14*, *PbUGT15*, *PbUGT56*, *PbUGT60*, and *PbUGT154* were all affected by salt stress (Fig. 8A). Generally, their expression levels exhibited a "rise-fall-rise" trend over time. *PbUGT14*, *PbUGT56*, *PbUGT60*, and *PbUGT154* reached their peak expression levels at 4 h, followed by an upward trend after a sudden decline. In contrast, the expression level of *PbUGT15* continuously decreased with time.

Under 10% PEG-6000 drought stress (Fig. 8B), except for *PbUGT15, PbUGT14, PbUGT56*, *PbUGT60*, and *PbUGT154* displayed peak expression levels at 6 h, followed by a sharp decrease and subsequent upward trend. *PbUGT15*, on the other hand, peaked at 4 h, exhibiting an overall 'rise-decline-rise' pattern. The expression levels of the five genes demonstrated significant variations under 1.9 M NaCl salt stress and 10% PEG-6000 drought stress, displaying both up-regulation and down-regulation. This indicates that these members of the gene family play crucial roles in the response of *P. bournei* to drought and salt stress.

Discussion

In the present investigation, a total of 156 *PbUGT* genes were identified in the genome of *Phoebe bournei*. The number of *PbUGT* genes in *P. bournei* exceeded that of *Arabidopsis thaliana* (122) and *Zea mays* (147), but was lower than *Oryza sativa* (180) and *Glycine max* (182). In comparison to other gene families, the *UGT* gene family demonstrated a higher abundance across all plant species. The analysis of species covariance revealed that *UGT* genes may have diverged earlier in the evolutionary process, exhibiting higher homology and closer linkage to closely related plants. Initial stages of species divergence exhibited some covariance, which gradually diminished as the species continued to diverge.

Analysis of the physicochemical properties of *P. bournei* identified that 145 genes (93%) had an average molecular mass greater than 30 KDa, with an average of 50 KDa, and contained more than 100 amino acids. This suggests that the majority of *PbUGT* genes encode high molecular weight proteins. Moreover, 136 (87%) of the proteins were found to be acidic, and 130 (83%) were hydrophilic. The instability coefficients of 110 proteins exceeded 40, indicating that most of the proteins regulated by *PbUGT* genes were both unstable and acidic. Subcellular localization analysis revealed an uneven distribution of *PbUGT* genes across organelles, with the majority located in mitochondria (88), followed by 43 in the cytoplasm, 15 in the nucleus, and the remaining in organelles such as vesicles and peroxisomes. The abundance of *PbUGT* genes in mitochondria may be related to the substantial energy requirements of glycosyltransferase, a process regulated by *UGT*. This finding provides a significant foundation and insight for further *UGT* gene research.

During natural growth, forest trees frequently encounter various abiotic stresses, including drought, low temperature, and salinity stresses, which impede their growth. These stresses can cause damage to plants, necessitating optimal water and fertilizer conditions for forest trees to achieve expected growth rates. However, this also results in greater nutrient consumption and longer recovery periods for the forest land. Genes play a vital role in enhancing the tolerance of forest trees to abiotic stresses [44]. Forest trees respond to stress signals by modulating the expression of various genes, which, in turn, regulate metabolic responses and control tree growth under stressful conditions.

The current experiment, in conjunction with previous studies and cis-element analysis, has demonstrated that *UGT* genes across species respond to diverse abiotic stresses, thereby enhancing plant defense mechanisms. For instance, maize mutants lacking the *ZmUGT2* gene failed to complete flavonol glycosylation *in vivo*, resulting in inhibited germination rates and suppressed stem and root growth under abiotic stress conditions, particularly salt and drought stresses [45]. In tea trees, abiotic stress suppressed the gene regulation of *CsUGT91Q2* responsible for

nerolidol glycol production, leading to impaired reactive oxygen species scavenging and reduced cold resistance, thereby increasing sensitivity to cold stress [46].

As transcription factors, *UGT* genes can activate the expression of stress-related genes, enhancing plant stress resistance and participating in the regulatory network governing adversity stress in forest trees. Phylogenetic analysis, transcriptome data, and RT-qPCR data confirmed that the five *PbUGT* genes can be rapidly induced and activated in response to abiotic stresses such as drought and salt stress, playing a significant role in the plant's response to these stresses.

Furthermore, transcriptome data analysis of *P. bournei* seedlings subjected to drought and salt stress revealed the involvement of all five *UGT* genes in abiotic stress responses. Their expression varied significantly with the duration of stress. The majority of genes exhibited up-regulated expression under drought and salt stress, displaying a 'rising-declining-rising' expression pattern under both 1.9 M NaCl salt stress and 10% PEG-6000 drought stress. Particularly, under 1.9 M NaCl salt stress, all five *UGT* genes exhibited a high expression pattern after only 4 h. These findings suggest that *PbUGT* genes play a positive regulatory role in abiotic stress response. This provides new insights for *P. bournei* species selection and stress regulation studies, aiming to expand the planting area of *P. bournei* and maximize its growth potential under stressful conditions.

Conclusion

The *UGT* gene family holds considerable significance in plant growth and development. In the case of *Phoebe bournei*, a comprehensive set of 156 *PbUGT* genes were successfully identified. Evolutionary analysis facilitated the classification of *P. bournei UGT* genes into fourteen distinct subfamilies. Notably, genes within the same subfamily often exhibited similar exon-intron structures, motifs, and protein configurations. The expression patterns of *PbUGT* genes demonstrated noteworthy variations under salt and drought stresses. Our experimental data substantiated the significant differences exhibited by *UGT* genes in response to diverse stresses, thereby emphasizing their crucial role in stress adaptation. This study serves as a foundational exploration into the functional attributes of the *PbUGT* gene, offering a theoretical framework for fellow researchers interested in investigating the response of *PbUGT* genes under abiotic stress conditions. In light of mounting challenges posed by land salinization and drought, further research on enhancing *P. bournei* species could contribute to expanding the cultivation area dedicated to this species.

Methods

Identification, Chromosomal Localization, and Analysis of Protein Physicochemical Properties of Phoebe bournei UGT Transcription Factor Family Members

The whole genome data and annotation files of *Phoebe bournei* were acquired from the China National GeneBank DataBase (CNGBdb) with lookup number CNP0002030 (https://db.cngb.org/). The Pfam website was utilized to obtain the *UGT* structural domain (query code PF00201) for screening conserved sequences (E-value < 1e-5) in all *P. bournei* protein sequences using Hmmer 3.3.2 software (http://hmmer.org/) [23]. Furthermore, the CDD function of the NCBI website (https://www.ncbi.nlm.nih.gov/cdd) and SMART online software (https://smart.embl.de/) [24] were employed to analyze and integrate the obtained protein sequences. Erroneous sequences were manually removed, ensuring the presence of UDP-glycosyltransferase structural domains. Ultimately, 156 sequences containing the complete *UGT* structural domain were selected.

Based on the whole genome data and annotation files of *P. bournei*, specific location information for *UGT* genes on each chromosome was obtained. Subsequently, Mapchart 2.32 software [27] was employed to map the location information of each *PbUGT* gene onto the 12 *P. bournei* chromosomes. The ExPASy online program (https://web.expasy.org/protparam) [25] was utilized to predict and calculate the molecular weight, number of amino acids, theoretical PI (isoelectric point), number of positive (negative) residues, instability coefficient, lipolysis coefficient, and total average hydrophilicity of each UGT protein in *P. bournei*. The subcellular localization of each member of the *P. bournei UGT* gene family was predicted using the WoLFPSORT website (https://wolfpsort.hgc.jp) [26].

Gene Structure, Protein Conserved Motifs, and Exon-Intron Analysis of Phoebe bournei UGT Genes

The annotation file of the *P. bournei* genome and the amino acid sequences of the 156 *UGT* genes were used to construct a phylogenetic evolutionary tree for *PbUGT* genes using the One Step Build an ML Tree function in TBtools v1.120 [28]. The MEME 5.5.2 online software (http://meme-suite.org/tools/meme) [29] was employed to analyze the conserved motifs of *PbUGT* genes with a motif set of 10. The GSDS2.0 online tool (http://gsds.cbi.pku.edu.cn/) [30] was utilized to map the gene structure, including UTRs, exons, and introns of each UGT gene. The Batch CD-search function of the NCBI website (https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi) was used to query the structural domains by submitting the gene sequences. The Gene Structure View function in TBtools [28] facilitated the visualization and mapping of phylogenetic relationships, conserved motif composition, structural domains, and exon-intron gene structures for each *PbUGT* gene. The resulting maps were subsequently combined.

Phylogenetic Analysis of Phoebe bournei UGT Genes and Construction of the Phylogenetic Tree

Nucleotide and amino acid sequences of 122 *Arabidopsis thaliana UGT* genes were obtained from the *A. thaliana* glycosyltransferase family 1 information website (http://www.p450.kvl.dk/UGT.shtml) [31] [32] [33] as a reference. Multi-protein sequence alignment was performed using MUSCLE [34] on the combined sequences of 156 UGT genes from *P. bournei*. Phylogenetic trees of related proteins were constructed using the maximum likelihood method in the MEGA11 program with a Bootstrap value set to 1000 and default parameters [35]. The constructed phylogenetic trees were further organized and embellished using the iTol online website (https://itol.embl.de/) /) [36].

Analysis of Tissue Expression Profiles of PbUGT Genes

The sequence number PRJNA628065 of *P. bournei* was searched on the European Bioinformatics Centre repository EBI (https://www.ebi.ac.uk). Transcriptome expression data of *P. bournei* in five different tissues were downloaded [19]. RNA-seq analysis was conducted on the data to investigate the differential expression of the 156 *PbUGT* genes. TBtools v1.120 software was used to algebraically transform the FPKM values from the analyzed data into log2 values, which were then used to generate heat maps.

Covariance Analysis of UGT Family Homologous Genes in Phoebe bournei, Arabidopsis thaliana, Populus hirsutus, and Cinnamomum camphora

To elucidate the covariance patterns among *PbUGT* family homologous genes in *Arabidopsis thaliana*, *Populus hirsutus*, and *Cinnamomum camphora*, a genome-wide covariance analysis was conducted. This analysis integrated the chromosome-level genome data of *P. bournei* with the genome-wide data of *A. thaliana* [37], *P. hirsutus* [38], and *C. camphora* [39]. Synteny Visualization function in TBtools was employed to map the covariance across species. The whole-genome data of *A. thaliana* and *P. trichocarpa* from the Ensembl Plants database

(https://plants.ensembl.org/index.html) and *C. camphora* genome-wide data from the National Genomics Data Center (https://ngdc.cncb.ac.cn/) (login number: GWHBGXC00000000) were utilized.

Replication Analysis of PbUGT Gene Duplications on Chromosomes

Circos [40] was used to identify each *PbUGT* gene on individual chromosomes. The results were then visualized using the Multiple Covariance Scanning Toolkit MCScan [41] with default settings. A comparison within the entire *P. bournei* genome was performed based on amino acid sequence homology. The duplication events in each *PbUGT* gene were analyzed, and the results were visualized using Circos and Dual Synteny Plot in TBtools.

Enrichment Analysis of Stress, Hormone, and Light-Responsive Cis-Elements in the Promoter Region of PbUGT Genes

The GTF/GFF3 Sequences Extract function of TBtools was employed to obtain a 2000 bp segment upstream of the promoter region of each *PbUGT* gene based on the whole-genome sequencing sequence of *Phoebe bournei.* PlantCare (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [42], an online software, was used for prediction and screening of cis-acting elements associated with *PbUGT* genes. Subsequently, the data were categorized, organized, and analyzed, and the Heatmap function in TBtools was utilized to classify and visualize the number of cis-acting elements for each stress-responsive *PbUGT* gene.

Plant Material and Abiotic Stress Treatments in Phoebe bournei

One-year-old *P. bournei* seedlings from Fujian Academy of Forestry were selected as the plant material and incubated in artificial climate chambers under different treatments. The seedlings with consistent growth potential were divided into two groups: the control group and the stress treatment group. The treatment group comprised 30 plants, while the control group consisted of three plants. For each biological replicate, every two plants were included in the treatment group. Three sets of biological replicates were established for each time point. Drought stress was induced using a 10% PEG-6000 solution, while salt stress involved culturing *P. bournei* seedlings in a 1.9 M NaCl solution. The control group was maintained under normal moisture conditions. All stress treatments were conducted for 4, 6, 8, and 12 h. After the designated treatment periods, leaf samples from the *P. bournei* seedlings were collected, immediately frozen in liquid nitrogen, and stored for subsequent RNA extraction.

Real-time Quantitative Polymerase Chain Reaction (RT-qPCR) Analysis of PbUGT Genes

RNA extraction was performed using the HiPure Plant RNA Mini Kit (Magenbio, Canton, China), and cDNA synthesis was carried out using the PrimeScript RT reagent Kit (Perfect Real Time) kit (TaKaRa, Dalian, China). Based on the types of cis-acting elements present in the *PbUGT* gene family members, the genes with the highest number of elements associated with drought and salt stress response (*PbUGT14*, *PbUGT15*, *PbUGT56*, *PbUGT60*, and *PbUGT154*) were selected for RT-qPCR analysis. The primer sequences ofs quantitative RT-qPCR are shown in Table 2. Specific primers in the non-conserved regions of the target genes were designed using Primer 3.0 software and synthesized by Fuzhou Qingbaiwang Biotechnology Company. Real-time fluorescence quantitative analysis was performed using 1 μL of cDNA template, 10 μL of SYBR Premix Ex TaqTM II, 2 μL of specific primers, and 7 μL of ddH_2O. The reaction program consisted of 40 cycles: 95°C for 30 s, 95°C for 5 s, 60°C for 30 s, 95°C for 5 s, 60°C for 60 s, and 50°C for 30 s. The internal reference gene used was *PbEF1α* (GenBank No. KX682032) [43]. The expression of the target gene was calculated using the $2^{-\Delta\Delta Ct}$ method. GraphPad Prism 8.0 software was utilized

for statistical analysis, performing t-tests and generating graphs. A significance level of $P < 0.05$ was used to determine statistical significance.

Table 2
Primer sequences of quantitative RT-qPCR

Primer	Sequence (5'~3')
PbUGT14(*OF21457-F*)	TGCATGGGGCCTCAAAGG
PbUGT14(*OF21457-R*)	GGGCGAGCACTTCCAACT
PbUGT15(*OF04928-F*)	GTTTCCGCAGTGGGGTGA
PbUGT15(*OF04928-R*)	CGGTCCCCCAACAACCTC
PbUGT56(*OF23900-F*)	GTTCGTGGGGTGGAGTGG
PbUGT56(*OF23900-R*)	GGGTGGGCCAGAACTTCC
PbUGT60(*OF22991-F*)	TCGCCGCATCGAATCCTC
PbUGT60(*OF22991-R*)	AGGCGAGGGATTCAGGGT
PbUGT154(*OF17153-F*)	GCGGGCAGAGGTTCTTGT
PbUGT154(*OF17153-R*)	TTGCGACGCCCATGAGTT
Internal reference gene (*EF1α-F*)	CATTCAAGTATGCGTGGGT
Internal reference gene (*EF1α-R*)	ACGGTGACCAGGAGCA

Declarations

Acknowledgements

Not applicable.

Authors' contributions

SJC conceived the study, ZNZ and YZZ performed experiments, data analysis and manuscript writing. JSL worked on RT-qPCR. SJY and JXP performed on RNA isolation and HFG AB helped in the experiments. SJC and YZZ revised the manuscript. SPC supported the project with funds. All authors read and approved the final manuscript.

Funding

This research was jointly funded by Fujian Agriculture and Forestry University Forestry peak discipline construction project (71201800739 to Shijiang Cao) and the 7th Project of Forest Seeding Breaking in Fujian Province (LZKG-202208 to Shipin Chen).

Availability of data and materials

Publicly available datasets were analyzed in this study. This data can be found here: The P. bournei genome data presented in this study are openly available in CNGBdb at https://db.cngb.org/, accessed on 10 February 2023, CNP0002030. The A. thaliana and P. trichocarpa genome data presented in this study are openly available in

EnsemblPlents at https://db.cngb.org/, accessed on 10 February 2023. The C. camphora genome data presented in this study are openly available in NSTI at https://ngdc.cncb.ac.cn/, accessed on 10 February 2023. The P. bournei RNA seq data presented in this study are openly available in https://www.ncbi.nlm.nih.gov/bioproject/,accessed on 10 February 2023, PRJNA628065.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional afliations

References

1. Vogt T, Jones P. Glycosyltransferases in plant natural product synthesis: characterization of a supergene family. Trends Plant Sci. 2000;5(9):380–6.
2. Keegstra K, Raikhel N. Plant glycosyltransferases. Curr Opin Plant Biol. 2001;4(3):219–24.
3. J H, A H M. Multiple secondary plant product UDP-glucose glucosyltransferase genes expressed in cassava (Manihot esculenta Crantz) cotyledons. DNA Seq. 1994;5(1):41–9.
4. Kanzaki S, Kamikawa S, Ichihi A, Tanaka Y, Shimizu K, Koeda S. Isolation of UDP:flavonoid 3-O-glycosyltransferase (UFGT)-like Genes and Expression Analysis of Genes Associated with Anthocyanin Accumulation in Mango 'Irwin' skin. Hortic J. 2019;88(4):435–43.
5. Bo W, Shang-Hui J, Hong-Qun H, Yan-Guo S, Yan-Wen W, Ping H, Bing-Kai H. UGT87A2, an Arabidopsis glycosyltransferase, regulates flowering time via FLOWERING LOCUS C. The New phytologist,2012;194(3):666–75.
6. Paquette S, Møller BL, Bak S. On the origin of family 1 plant glycosyltransferases. Phytochemistry. 2003;62(3):399–413.
7. T BV, C PV. Phylogenomic analysis of UDP glycosyltransferase 1 multigene family in Linum usitatissimum identified genes with varied expression patterns. BMC Genomics. 2012;13(1):175.
8. Yi H, Dawood A, Xu Z, Yu Z, Lei W, Peng J, Hongxiang M. Genome-wide analysis of family-1 UDP glycosyltransferases (UGT) and identification of UGT genes for FHB resistance in wheat (Triticum aestivum L). BMC Plant Biol. 2018;18(1):67.
9. Yanjie L, Pan L, Yong W, Ruirui D, Huimin Y, Bingkai H. Genome-wide identification and phylogenetic analysis of Family-1 UDP glycosyltransferases in maize (Zea mays). Planta. 2014;239(6):1265–79.

10. Xianghui X, Quanwei L, Ruixian L, Juwu G, Wankui G, Aiying L, Qun G, Junwen L, Haihong S, Pengtao L. Genome-wide characterization of the UDP-glycosyltransferase gene family in upland cotton. 3 Biotech. 2019;9(12):453.
11. Boping W, Liuxiao G, Jie G, Yaying X, Hongru L, Xiangmei C, Bo Z, Kunsong C. Genome-Wide Identification, Expression Patterns, and Functional Analysis of UDP Glycosyltransferase Family in Peach (Prunus persica L. Batsch). Front Plant Sci. 2017;8:389.
12. Lorenzo C, Mickael M, Vadim G, Svetlana N, Stefan M. 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: for cell and molecular biology. 2012;69(6):1030–42.
13. Cao P, Bartley LE, Jung K, Ronald PC. Construction of a Rice Glycosyltransferase Phylogenomic Database and Identification of Rice-Diverged Glycosyltransferases. Mol Plant. 2008;1(5):858–77.
14. Wu B, Liu X, Xu K, Zhang B. Genome-wide characterization, evolution and expression profiling of UDP-glycosyltransferase family in pomelo (Citrus grandis) fruit. BMC Plant Biol. 2020;20(1):459.
15. Brigitte P, Franz B, Doris L, Tobias S, Rainer S, Rudolf K, Karl K, Josef G, Christian L, Gerhard A. Detoxification of the Fusarium mycotoxin deoxynivalenol by a UDP-glucosyltransferase from Arabidopsis thaliana. J Biol Chem. 2003;278(48):47905–14.
16. Xiwen L, Jie L, Puhua H, Fanan W, Hung-pin, Tsui. Henk van der Werff. Lauraceae. Flora of China. Volume 7. Beijing: Science Press & St; 2008.
17. Darong W, Zhengde Z. Preliminary study on structure and spatial distribution pattern of Phoebe bournei in Luo Boyan nature reserve in Fujian province. Scientia Silvae Sinicae. 2003;39(1):23–30.
18. ShiPin C, WeiHong S, YuanFang X, YuTing J, XueDie L, XingYu L, DiYang Z, ShuZhen J, Yu L, Bin L. The Phoebe genome sheds light on the evolution of magnoliids. Hortic Res. 2020;7(1):146.
19. Xiao H, Junhong Z, Shuang H, Li CS, Guanliang M, Minyan S, Yang W, Shengcai Z, Chengcheng L, Luhuan L. The chromosome-scale genome of Phoebe bournei reveals contrasting fates of terpene synthase (TPS)-a and TPS-b subfamilies. Plant Commun. 2022;3(6):100410.
20. Miao Z, Shengcai Z, Mengjie W, Zaikang T, Xiao H, Junhong Z, Longjun C. Identification of the PbWRKY gene family and its expression analysis under deficiency of phosphorus in Phoebe bournei. Scientia silvae sinicae. 2022;58(02):133–47.
21. Shuang H, Xiao H, Yibei L, Yuting Z, Junhong Z, Zaikang T. Identification of NF-Y gene family and expression analysis in response to drought stress in Phoebe bournei. J agricultural Biotechnol. 2022;30(06):1112–27.
22. Qianqing W, Yuting Z, Junhong Z, Hui L, Zaikang L. Identification of PLR gene family and expression of responsive hormones in Phoebe bournei. J Zhejiang A&F Univ. 2022;39(06):1173–82.
23. Jody DFR. R E S. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:29–37.
24. Ivica L, Tobias D, Peer B. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 2012;40:302–5.
25. Elisabeth G, Alexandre G, Christine H, Ivan I, Amos DAR. ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 2003;31(13):3784–8.
26. Paul H, Keun-Joon P, Takeshi O, Naoya F, Hajime H, Kenta JAC. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35:585–7.

27. E V R. MapChart: software for the graphical presentation of linkage maps and QTLs. J heredity. 2002;93(1):77–8.
28. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. Mol Plant. 2020;13(8):1194–202.
29. Mikael LBT, Martin BABF, Luca FEGC, Jingyuan C, W R, LW, S NW. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:202–8.
30. Bo H, Jinpu J, An-Yuan G, He Z, Jingchu L, Ge G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7.
31. I MP SOI. W HD, T I. The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. Pharmacogenetics. 1997;7(4):255–69.
32. Y L, S B, K LE, J BD. Phylogenetic analysis of the UDP-glycosyltransferase multigene family of Arabidopsis thaliana. J Biol Chem. 2001;276(6):4338–43.
33. J R, Y L, E L, J BD. Higher plant glycosyltransferases. Genome Biol. 2001;2(2):reviews3004.
34. C E R. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32(5):1792–7.
35. Koichiro T, Glen S, Sudhir K. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Mol Biol Evol. 2021;38(7):2725–9.
36. Ivica L, Peer B. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49(W1):293–6.
37. Chia-Yi C, Vivek K, P CA, Françoise T, Seth S. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. The Plant journal: for cell and molecular biology. 2017;89(4):789–804.
38. Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A. The Genome of Black Cottonwood, Populus trichocarpa (Torr. &Gray). Science. 2006;313(5793):1596–604.
39. Sun W, Xiang S, Zhang Q, Xiao L, Zhang D, Zhang P, Chen D, Hao Y, Liu D, Le Ding. The camphor tree genome enhances the understanding of magnoliid evolution. J Genet Genomics. 2022;49(03):249–53.
40. Martin K, Jacqueline S, Inanç B, Joseph C, Randy G, Doug H. A MM. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639–45.
41. Yupeng W, Haibao T, D DJ, Xu T, Jingping L, Xiyin W, Tae-ho L, Huizhe J, Barry M, Hui G. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.
42. Magali L, Patrice D, Gert T, Kathleen M, Yves M, Van de Yves P, Pierre R, Stephane R. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325–7.
43. Zhang J, Zhu Y, Pan Y, Huang H, Li C, Li G, Tong Z. Transcriptomic profiling and identification of candidate genes in two Phoebe bournei ecotypes with contrasting cold stress responses. Trees. 2018;32(5):1315–33.
44. Bressan R, Bohnert H, Zhu J. Abiotic Stress Tolerance: From Gene Discovery in Model Organisms to Crop Improvement. Mol Plant. 2009;2(1):1–2.
45. Yan-Jie L, Pan L, Ting W, Feng-Ju Z, Xu-Xu H, Bing-Kai H. The maize secondary metabolism glycosyltransferase UFGT2 modifies flavonols and contributes to plant acclimation to abiotic stresses. Ann Botany. 2018;122(7):1203–17.

46. Mingyue Z, Na Z, Ting G, Jieyang J, Tingting J, Jingming W, Yi W, Xiaochun W, Wilfried S, Chuankui S. Sesquiterpene glucosylation mediated by glucosyltransferase UGT91Q2 is involved in the modulation of cold stress tolerance in tea plants. New Phytol. 2020;226(2):362–72.

Figures

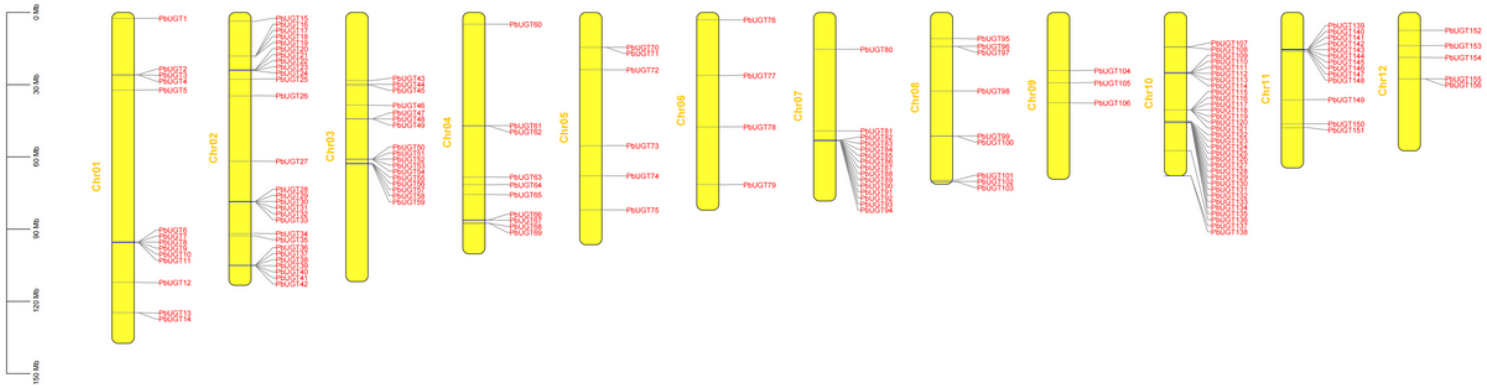


Figure 1

Chromosomal distribution and duplication of *PbUGT*genes.

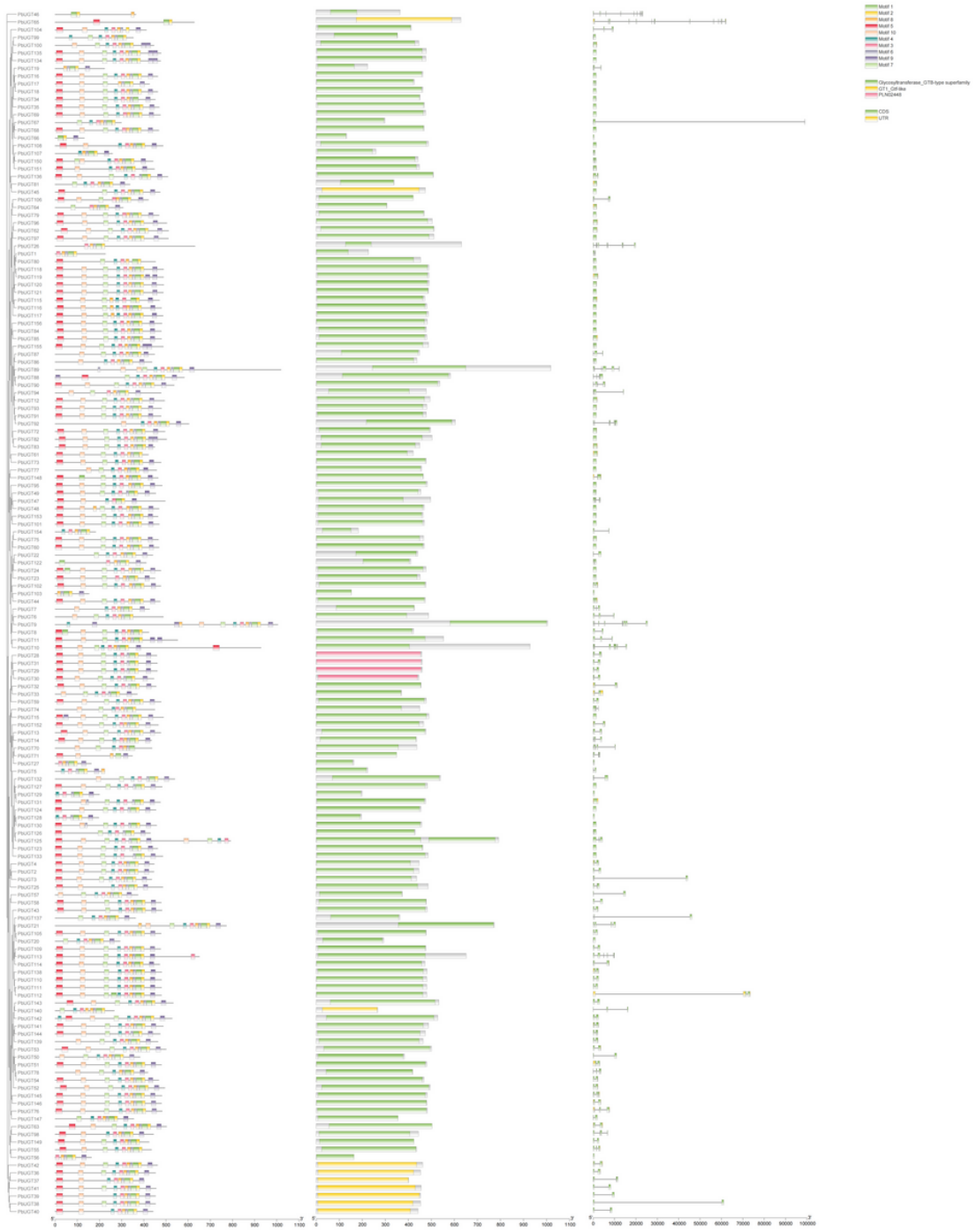


Figure 2

Phylogenetic relationship, conserved motif, domain and exon-intron structures of *PbUGT*gene family.



Figure 3

Phylogenetic tree of UGT proteins of *Phoebe bournei* and *Arabidopsis thaliana.*

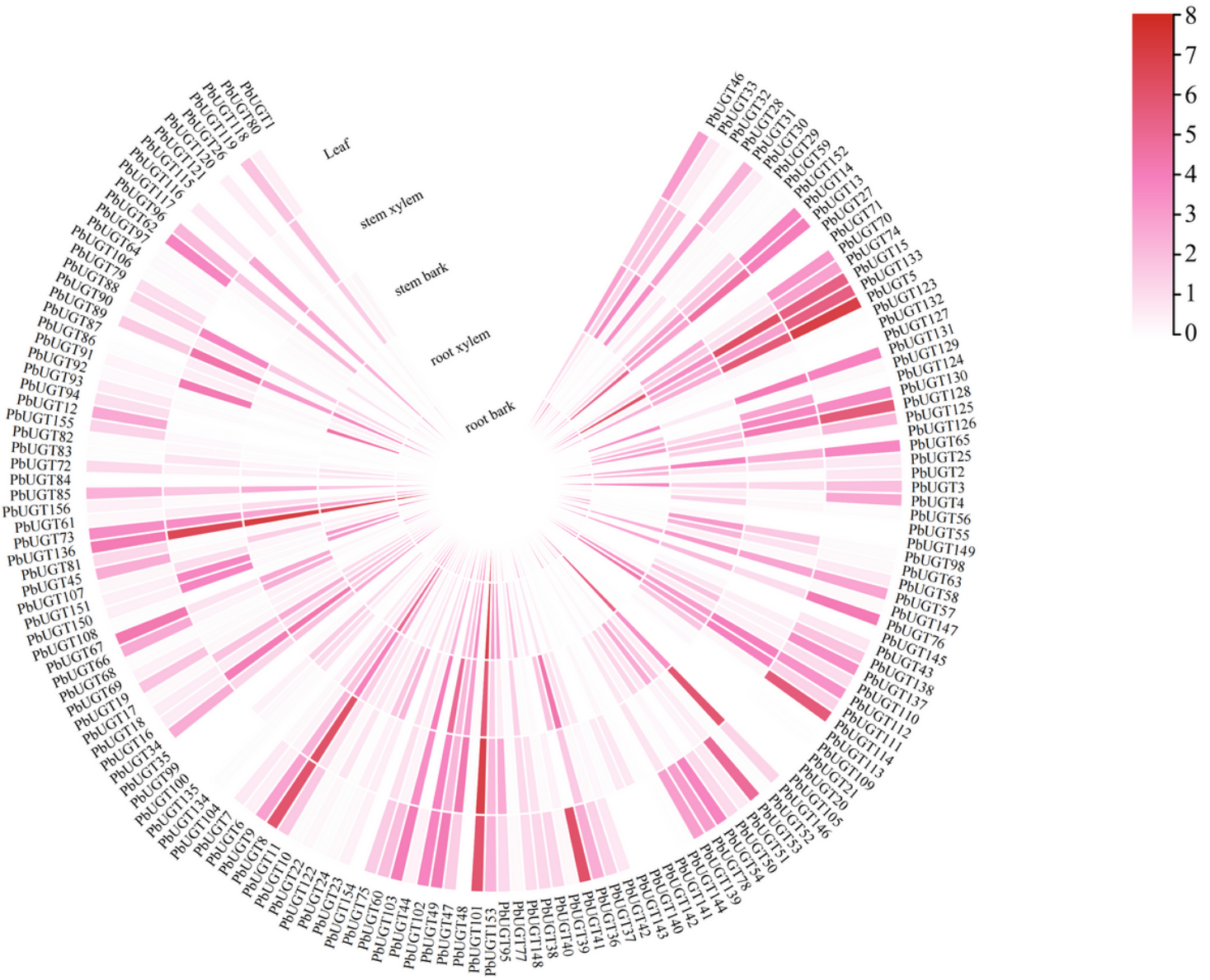


Figure 4

Expression heatmap of the *PbUGT* family in different tissues. Differences in gene expression changes are shown by color, with darker colors indicating higher expression.

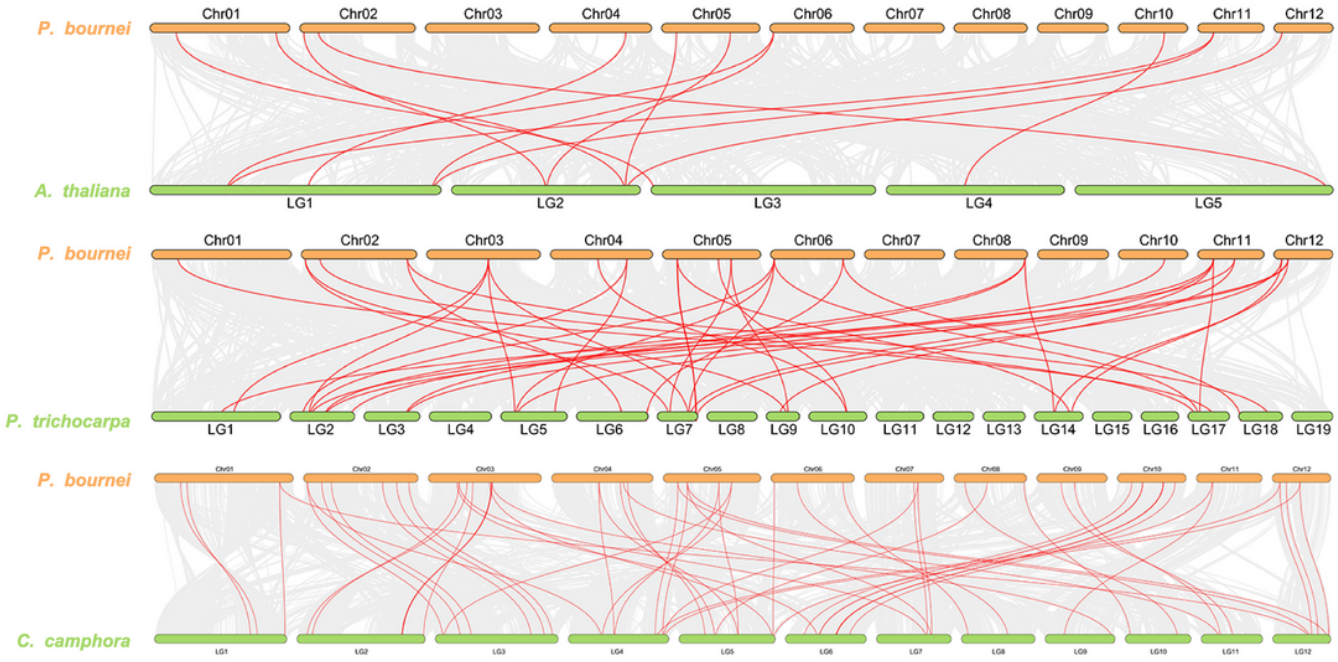


Figure 5

Synteny analysis of *UGT* genes among *Phoebe bournei*, *Arabidopsis thaliana*, *Populus trichocarpa* and *Cinnamomum camphora.* The two rings in the middle represent the gene density of each chromosome, the gray lines represent the collinearity, and the red lines represent the highlighted collinearity of *UGT* genes

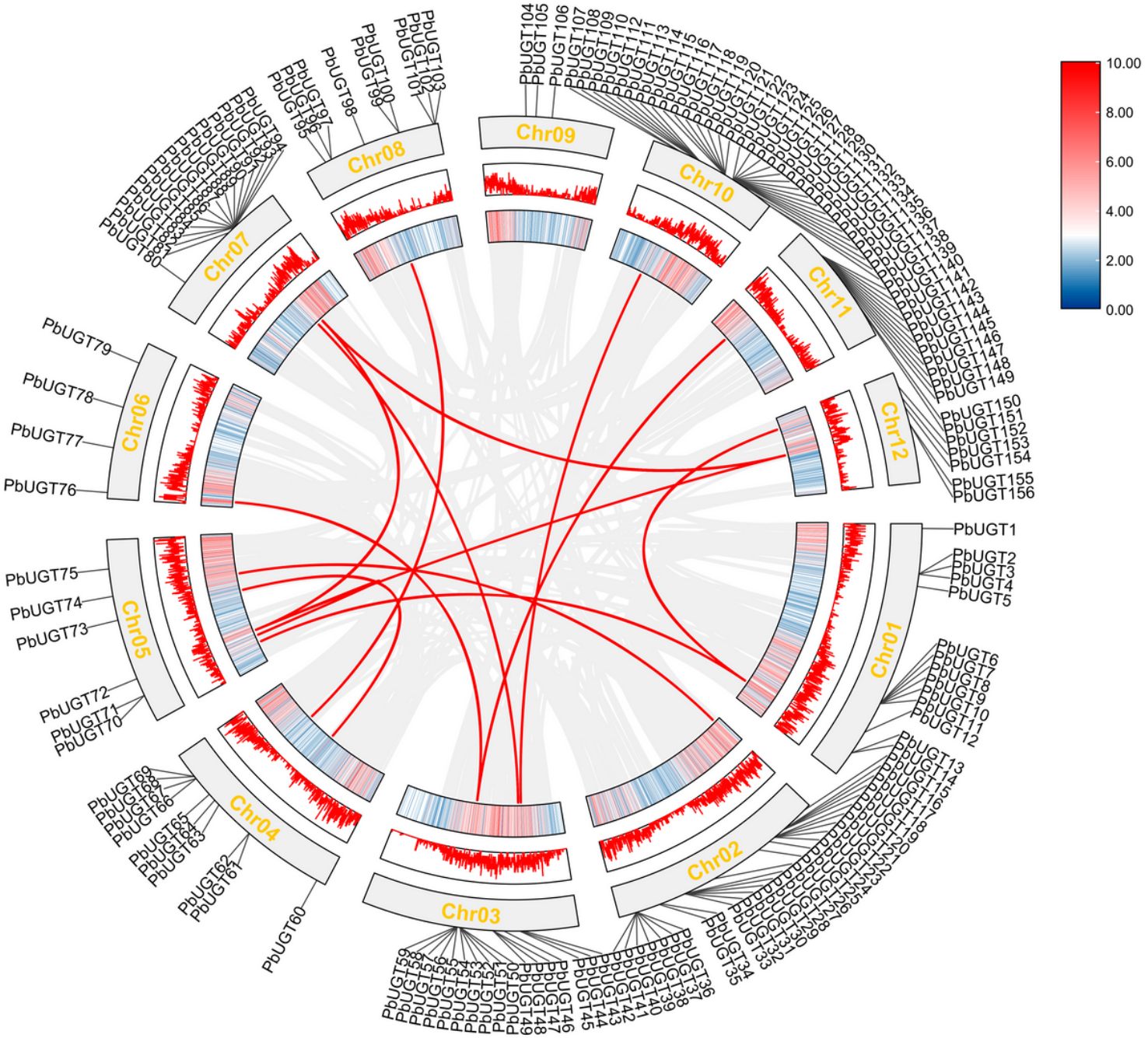


Figure 6

Distribution and collinearity analysis of *PbUGT*gene. The two rings in the middle represent the gene density of each chromosome, the gray lines represent the collinearity, and the red lines represent the highlighted collinearity of *UGT* genes

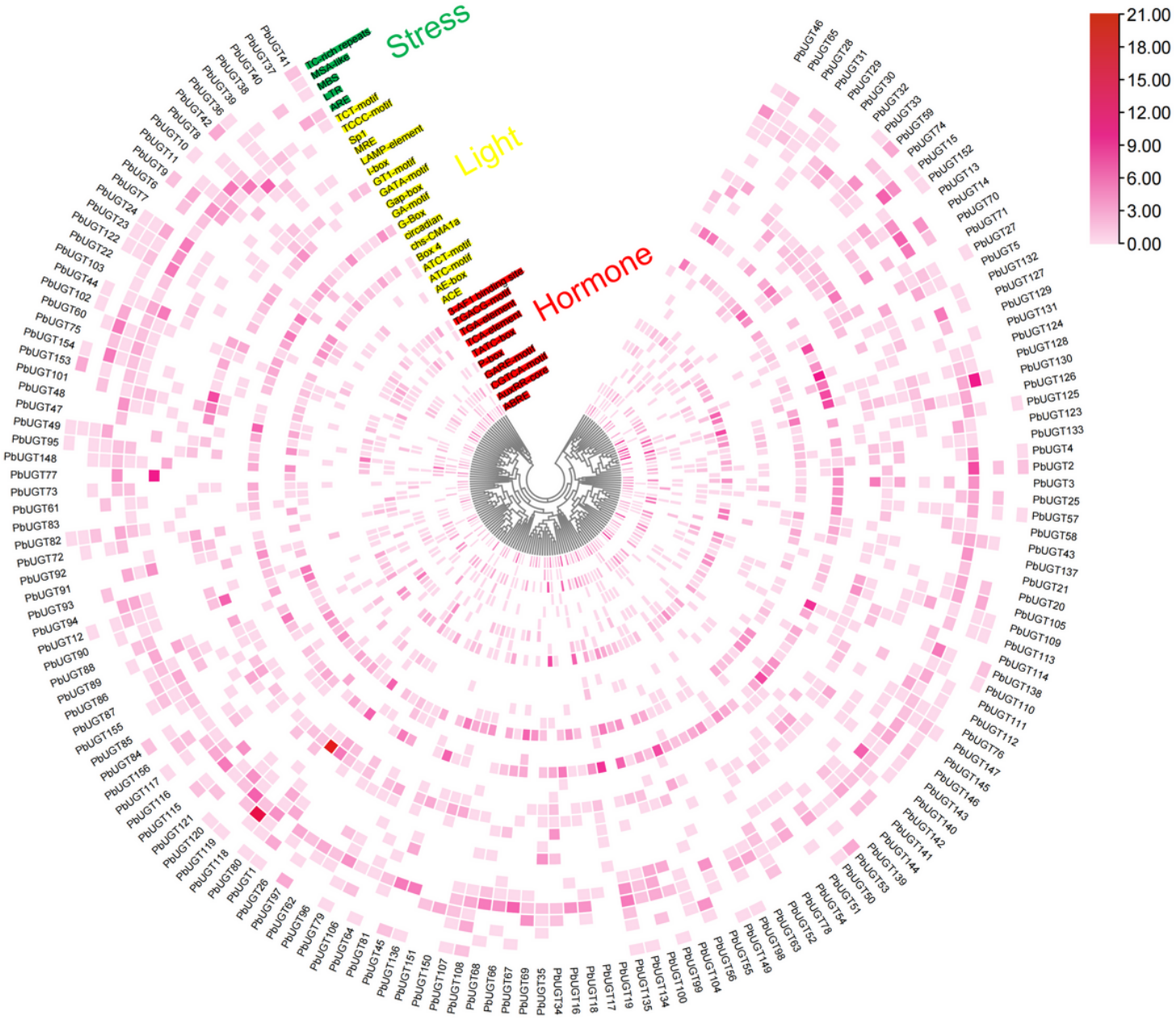


Figure 7

Enrichment analysis of stress, hormone and light-responsive *cis*-elements in the promoter regions of *PbUGT*gene family. Differences in the number of *cis*-elements are indicated by color, with darker colors indicating more *cis*-elements.

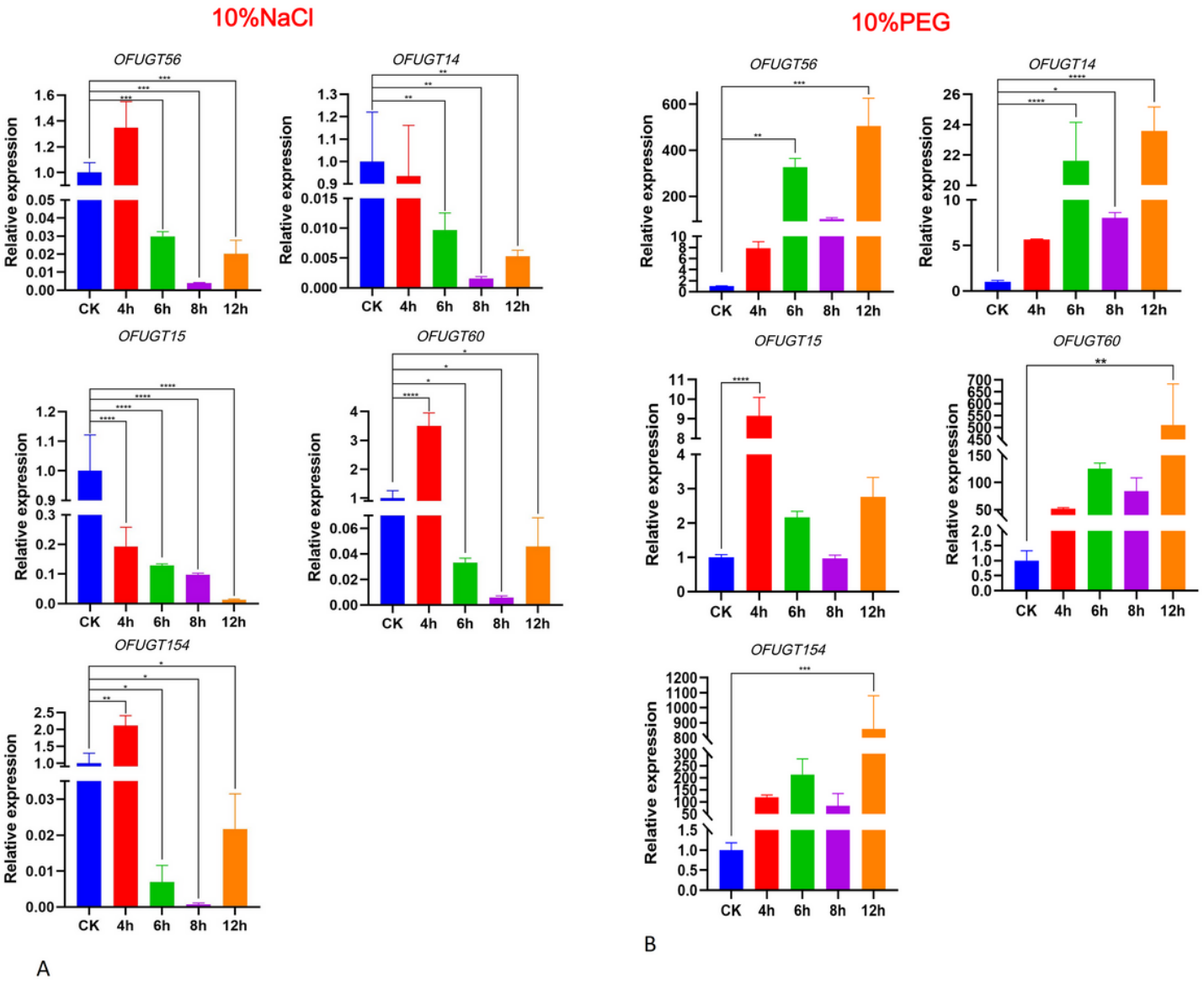


Figure 8

Expression profile of *PbUGT* genes responding to salt and drought stresses tested by RT-qPCR. (A) The relative gene expression levels under salt (1.9 M NaCl) treatments for the same periods (0, 4, 6, 8,12h). Control seedlings were treated with distilled water. (B) The relative gene expression levels under drought (10% PEG) treatments for the same periods (0, 4, 6, 8,12h). Control seedlings were treated with distilled water. (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0005$, **** $p < 0.0001$).