Systematic Analysis of GT1 Family Genes and Their Regulation in Anthocyanin Metabolism in Red Maple

Hao Xu
Anhui Finance & Trade Vocational College

Qing Zhu
Anhui Finance & Trade Vocational College

Xiaoyu Lu (maggie890407@126.com)
Anhui Finance & Trade Vocational College

Keywords: glycosyltransferases, GT1, red maple, anthocyanin, regulation

Posted Date: November 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2123334/v2

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

Glycosyltransferases (GTs) have a crucial role in the glycosylation of secondary metabolites, detoxification of endogenous or exogenous substances, body defense, and hormone regulation in plants. The GT1 family has the largest number of characterized enzymes in 111 GT families that are known for their excellent glycosylation capacities toward numerous valued small molecules. The colorful foliage of red maple is a vital agronomic trait, and studies have confirmed that massive anthocyanin accumulation led to the redness of leaves. In red maple, glycosylation is the essential step of anthocyanin biosynthesis and the prerequisite of further modifications, which usually enhances stability. Here, a genome-wide characterization analysis of the GT1 family in red maple was performed. A total of 560 GT genes were identified in the red maple genome; among them, 122 members belonged to the GT1 family. All these members were unevenly distributed across 19 chromosomes, with most located in the chloroplast. These GT1 genes had 1–16 exons. Most 122 GT1 proteins in red maple contained GT-GTB-type domain and GT1-Gtf-like domain. In total, 18 GT1 proteins might have played pivotal evolutionary roles in red maple. The network analysis revealed that the regulatory effect of GT1 family genes on anthocyanin in red maple leaves could be divided into direct and indirect regulation. The study results not only clarified the roles of the GT1 family in red maple but also laid a cornerstone for further functional analysis of this gene family in Acer plants.

Introduction

Glycosylation uses glycosyltransferases (GTs) to form specific glycosidic bonds between sugar and natural products, thereby synthesizing glycosidic compounds. This is the most extensive chemical reaction in nature (Meech et al., 2019). Through glycosylation modification, the biological activity and stability of aglycones increases (Janetzko, Trauger, Lazarus, & Walker, 2016). GTs can catalyze the formation of glycosidic bonds between specific sugars (donors) and receptors by using catalytic substrates such as sugars, proteins, lipids, and other small molecules (Ramírez et al., 2018). The most common sugar donor is activated nucleotide sugar, and the less common is phosphate-linked sugar (Zhang, Zhang, Zhang, Wang, & Wu, 2020). GTs are a class of highly differentiated, multi-member metabolic enzymes encoded by the multigene transferase family.

The Carbohydrate-Active EnZymes (CAZy) database (http://www.cazy.org) provides an amino acid sequence-based GT classification and has become the standard for classifying carbohydrate-active enzymes (Drula et al., 2022). Based on the similarity of GT sequences, the specificity of catalytic substrates, and the stereochemical structure of catalytic products, the CAZy has divided GTs into 111 families (Coutinho, Deleruy, Davies, & Henrissat, 2003) (GT1–GT114, and from them, GT36, GT46, and GT86 have now been removed). GT2, GT4, GT51, GT9, and GT1 are families with the largest number of GTs (301318, 232316, 74752, 38459, and 33106 GTs, respectively). The GT1 family has the largest number of characterized enzymes (330) in 111 GT families. They are known for their excellent glycosylation capacities toward numerous valued small molecules. GT1 enzymes in the plant kingdom also glycosylate numerous low-molecular-weight biologically active natural products, such as flavonoids,
benzophenones, terpenes, and steroids, that regulate the stability, solubility, and biological activity of aglycones and regulate plant hormones or exogenous biological detoxification (Bolam et al., 2007; Mulichak, Lu, Losey, Walsh, & Garavito, 2004).

Red maple, also referred to as *Acer rubrum* L., possesses a straight and tall stem and gorgeous leaf color and is native to the east coast of United States. It is among the most ornamental and prevalent species in the northern United States and some regions of Canada. Red maple was introduced to China before 1984 and received increased domestic attention around 1990. Great progress has been made in the research of physiological and ecological characteristics (Alexander & Arthur, 2010; Kalubia, Mehes Smith, & Omri, 2016), stress resistance mechanism (Kim, Im, & Nkongolo, 2016; Nkongolo, Theriault, & Michael, 2018), functional natural products (Geoffroy, Fortin, & Stevanovic, 2017), and leaf pigment composition (Schmitzer, Osterc, Veberic, & Stampar, 2009) of red maple. Red maple has broad market prospects, and therefore, its genetic improvement has become a research hotspot. Because of the completion of its whole-genome sequencing, basic biological research and breeding of red maple has been developed at the genetic level.

The colorful foliage of red maple, with vivid hues of green and red across different seasons, is one of its most crucial agronomic traits. Using technical methods involving the combination of transcriptomes and metabolomes, we confirmed that the massive accumulation of anthocyanins (particularly cyanidin derivatives) results in the redness of maple leaves (Z. Chen et al., 2019). In red maple, glycosylation is the essential step of anthocyanin biosynthesis and the prerequisite of further modifications, which typically enhances stability. Some GT1 family members are involved in anthocyanin glycosylation and metabolism (Rahimi et al., 2019).

We here systematically studied the GT1 family of red maple for the first time. In total, 122 GT1 family members in red maple were identified, and their gene structure, chromosomal location, gene replication events, collinearity, and phylogenetic tree were analyzed. Moreover, the correlated network analysis between GT1 family members and transcriptomic/metabolomic data revealed that GT1 genes play a role in anthocyanin biosynthesis in red maple. This genome-wide analysis of the GT1 family in red maple will provide a reference for the functional characteristics of genes belonging to this family in red maple.

**Materials And Methods**

**Plant materials**

All plant materials were collected from the Maple Research Base in Shucheng County, in Anhui Province (117°4′12″E, 31°28′48″N). From mid-September to mid-November in 2020, we collected good-quality green and red leaves for RNA isolation. The samples were stored at −80°C in a cryogenic refrigerator as described in a previous study (Lu et al., 2021).

**Database Search For Gts In Red Maple**
The proteins and their sequences of the *Arabidopsis* GT family were obtained from CAZy and NCBI databases (https://www.ncbi.nlm.nih.gov/), respectively. The sequences from these databases were queried during search against red maple genome databases with the BLASTP program. In total, 560 GT genes were identified from the red maple genome and named based on the nomenclature of *Arabidopsis* GT genes and their chromosomal locations.

**Analyses Of Chromosomal Localization, Gene Duplication, And Collinearity**

The TBtools software (C. Chen et al., 2020) was used to map the chromosomal location of GT1 family genes in red maple, according to the General Feature Format (GFF) files. Gene duplication in the GT1 family was analyzed and visualized using TBtools. The syntenic relationships between GT1 family members in red maple and other plants were displayed using the One Step MCScanX program of TBtools.

**Phylogenetics, Intron–exon Structures, Motif Composition**

MEGA7.0 software (Kumar, Stecher, & Tamura, 2016) was applied to prepare neighbor-joining (NJ) phylogenetic trees of GT1 family members in red maple and *Arabidopsis* with 100 bootstrap replicates. Exon and intro positions of these members were generated using TBtools according to GFF files of the red maple genome. Multiple Em for Motif Elicitation (MEME) was used to identify the conserved motif in the protein sequence of GT1 family members in red maple.

**Correlation Analysis**

The fragments per kilobase of the exon model per million mapped fragments (FPKM) value of each gene in transcriptomic data and the integral quantitative value of each metabolite in metabolomic data were collected as previously described. Spearman's correlation analysis was conducted using GraphPad Prism (8.0) software. Correlations with |R| > 0.9 and P < 0.05 were considered strong. Cytoscape software was employed to visualize the network of red maple GT1 family genes and anthocyanin derivatives.

**Results**

**Genome-wide identification and expression level analysis of GT gene families in red maple**

The GT family proteins in *Arabidopsis* and their sequences were obtained from CAZy and NCBI databases, respectively. BLASTP searches were performed to identify GT gene family members of red maple by using GT protein sequences of *Arabidopsis* as queries. We identified 560 GT genes from the red maple genome and named them based on the nomenclature of *Arabidopsis* GT genes and their
Volcano plots were established to represent the overall distribution of GT genes in red maple and illustrate the relationships between the P-value and log2 (fold change) (Fig. 1). In the red leaves (RL) versus green leaves (GL) group, 262 GT genes were upregulated, whereas 277 GT genes were downregulated. In the RL versus yellow leaves (YL) group, 296 GT genes were upregulated and 236 GT genes were downregulated. In the YL versus GL group, 233 GT genes were upregulated and 295 GT genes were downregulated. Based on the nucleotide diphospho-sugar, nucleotide monophospho-sugars, sugar phosphates, and related proteins, the red maple GT proteins were differentiated into sequence-based family classes (Table S1). GT43, GT57, GT59, and GT90 families were the smallest GT classes as they consisted of only one member. Most red maple GT genes encoded GT1 family proteins, which had 122 members.

**Analysis of subcellular location, chromosome distribution, and gene duplication modes of GT1 genes in red maple**

In the GT1 family of red maple (Table S2), the amino acid sequence length and molecular weight ranged from 155aa (ArUGT71C5) to 828aa (ArSGT2) and from 17836.42 to 91134.2 Da, respectively. The pl value of these GT1 family proteins in red maple ranged from 4.58 (ArUGT71C5) to 9.66 (ArALG14-1, ArALG14-2, and ArALG14-3). The prediction of subcellular location showed that most GT1 family proteins in red maple (59.03%) were located in the chloroplast; only one protein was located in mitochondria (ArUGT74F2-1) and the peroxisome (ArUGT85A1-1).

Figure 2 presents the result of chromosome location analysis of GT1 genes in red maple. In the red maple genome, 122 GT1 family genes were unevenly distributed across 19 chromosomes. Chromosome 24 contained the largest number of GT1 genes in red maple (24 genes), while only one gene each was located on chromosome 4, chromosome 8, and chromosome 26. GT1 genes in red maple were duplicated in four ways: segmental duplication, tandem duplication, proximal duplication, and dispersed duplication (Fig. S1, Table S2). The number of segment duplications were higher than that of tandem duplications.

**Analysis Of The Phylogenetic Relationship And Expression Level Of Gt1 Family Genes In Red Maple**

A NJ tree was constructed to investigate the phylogenetic relationships of GT1 genes by using their amino acid sequences in both red maple and *Arabidopsis*. Figure 3A shows that 122 GT1 genes belong to subgroups 1 (19 members), 2 (26 members), 3 (18 members), 4 (15 members), 5 (14 members), 6 (16 members), and 7 (14 members). A heatmap based on the aforementioned subgroups was plotted (Fig. 3B) to present the expression level of GT1 genes in red maple. The expression of 84.21% of subgroup 1 genes was the lowest in YL, while that of 83.33% of subgroup 2 genes was the lowest in RL. By contrast, the expression of 78.57% of subgroup 6 genes was the highest in RL among the three colored leaves.
Structural Analysis Of Gt1 Family Proteins In Red Maple

Figure 4 presents the motif and domain composition of GT1 proteins in red maple. All GT1 proteins in red maple contained motif 7, whereas most contained motif 2 and motif 4 (Fig. 4A). Interestingly, some motifs were presented only in certain GT1 subgroups. For example, motif 9 was discovered in subgroup 1, whereas all subgroup 2 members contained motif 5. These results suggest that the protein motifs were closely related to their subgroups. A protein molecule transcribed contains multiple regions with specific structures and different functions that are called domains. A domain is the basic unit of protein function, which is jointly determined by the multiple domains contained. Figure 4B shows that most 122 GT1 proteins in red maple contain the GT-GTB-type domain and GT1-Gtf-like domain. Four ArALG proteins contained the Alg14 domain, which is a UDP-N-acetylglucosaminyltransferase subunit. The exon arrangement of GT1 genes in red maple was analyzed in Fig. 4C. These GT1 genes had 1–16 exons (Table S3).

Evolutionary Analysis Of Gt1 Genes In Red Maple And Several Other Species

The collinearity analysis was performed to determine the evolutionary relationship between red maple and Citrus sinesis, Theobroma cacao, Arabidopsis thaliana, Vitis vinifera, Populus trichocarpa, and Acer yangbiense (Fig. 5). Ultimately, 41 collinear gene pairs between red maple and C. sinesis, 58 between red maple and T. cacao, 22 between red maple and A. thaliana, 16 between red maple and V. vinifera, 71 between red maple and P. trichocarpa, and 61 between red maple and A. yangbiense were identified. The results suggested that the gene pairs in all of the aforementioned plants, including 5 pairs (ArSIL1-5, ArUGT80B1-2, ArUGT80B1-3, ArALG14-4, and ArUGT84A2), 6 pairs (ArUGT71B7, ArUGT71D1-1, and ArUGT80B1-4), 7 pairs (ArUGT73B3-2), 8 pairs (ArUGT76E9 and ArUGT85A2-18), 9 pairs (ArUGT72E-5), 10 pairs (ArUGT85A2-14), 11 pairs (ArUGT73B3-3), 12 pairs (ArUGT85A2-2), 13 pairs (ArUGT85A2-3), 14 pairs (ArUGT86A1), and 15 pairs (ArUGT85A2-4), might have played pivotal evolutionary roles.

Network Analysis Of Gt1 Genes And Anthocyanin Derivatives In Red Maple

The Spearman test of these genes and anthocyanin derivatives was conducted to understand the interactive regulation of GT1 genes for anthocyanin derivatives in red maple. Correlations with $P < 0.05$ and $|r| > 0.9$ were strong. Based on the results (Table S4), the interactive network between GT1 genes and anthocyanin derivatives was established using Cytoscape software.

Figure 6 depicts that 41 GT1 genes and 9 anthocyanin derivatives were strongly correlated. Only two members of Subgroup 3 and Subgroup 7 were strongly correlated with four and five anthocyanin derivatives, respectively, while four members of Subgroup 5 were strongly correlated with four
anthocyanin derivatives. Notably, 28 strong correlations were observed between the eight members of Subgroup 4 and six anthocyanin derivatives.

Discussion

Identification of GT gene families in red maple

Our previous studies have confirmed that red maple leaves turn red in autumn because of anthocyanin accumulation. However, water-soluble anthocyanin monomers are unstable in plants. Anthocyanin glycosylation is essential for anthocyanin biosynthesis and the prerequisite of further modifications, which generally enhance stability (Zhao et al., 2014). Figure 6 and Table S4 jointly describe the regulatory relationship between GT1 family genes and anthocyanin derivatives in red maple leaves, which might be categorized as direct and indirect regulation.

Direct regulation

UGT78D2 specifically glycosylates the 3-position of the flavonoid C-ring (Lim, Ashford, Hou, Jackson, & Bowles, 2004) (Fig. 7A). Anthocyanin such as cyanidin is also the accepted substrate (Tohge et al., 2005). Under the catalysis of UGT78D2, a flavonol and a component of UDP-α-D-glucose react to produce a type of flavonol 3-O-β-D-glucoside and a component of \( H^+ \) and UDP. Figure 6 shows that ArUGT78D2-1, ArUGT78D2-2, and ArUGT78D2-4 have a strong correlation with Cy-acetylglucoside. Analysis of the chemical structure formula of Cy-acetylglucoside revealed that it was a variant of flavonol 3-O-β-D-glucoside, which acetyl- replaced cis-hydroxyl- (Fig. S2). -WAPQ- and -HCGWNSVLE- at 388–364 bp in AtUGT78D2 was considered the binding site of UDP-D-glucose. By aligning the protein sequences (Fig. 7B), ArUGT78D2-2 and ArUGT78D2-4 had the same binding site as AtUGT78D2, while the sequence of ArUGT78D2-1 at the same position was similar to that of AtUGT78D2. Thus, the similar chemical structure between Cy-acetylglucoside and flavonol 3-O-β-D-glucoside, and the similar binding sites between ArUGT78D2-1, ArUGT78D2-2, ArUGT78D2-4, and AtUGT78D2 might be the reason for the direct regulation of Cy-acetylglucoside by ArUGT78D2 genes.

Indirect regulation

UGT73B2 catalyzed the glycosylation of flavonoids from UDP-glucose, which also glycosylates with activity kaempferol and quercetin at the 3-O position (Willits et al., 2004) (Fig. 8A). During anthocyanin biosynthesis, the metabolism of its precursors, namely didrokaempferol, dihydroquercetin, and dihydromyricetin, is divided into two branches (Fig. 8B). One of which involved the final production of pelargonidin, cyanidin, and delphinidin, respectively; the other branch involved the production of kaempferol, quercetin, and myricetin, respectively, the UDP-glucose binding sites of UGT73B2 in Arabidopsis thaliana were reported to be -APQ- in 355–357 bp and -HCGWNSVLE- in 372–380 bp. Aligning analysis showed that ArUGT73B2 proteins (Fig. 8C) had the same binding site as AtUGT73B2. R values of ArUGT73B2-1, ArUGT73B2-2, and ArUGT73B2-5 with anthocyanin derivatives were positive.
(Table S4), indicating that these genes positively regulate anthocyanin biosynthesis in red maple. Knocking out one allele of VvbZIP36 in grapevine promoted anthocyanin accumulation but inhibited quercetin-3-O-rhamnoside synthesis (Tu et al., 2022). In this study, the positive regulation of anthocyanin derivatives by three ArUGT73B2 genes confirmed this result. In the kaempferol and quercetin branch, the increased consumption of kaempferol and quercetin by ArUGT73B2 genes might decrease kaempferol and quercetin accumulation, so that the metabolic flow was biased toward anthocyanin synthesis and the synthesis of anthocyanin derivatives increased.

In addition to regulating anthocyanin synthesis through metabolites in the pathway, GTs regulate anthocyanin synthesis through hormones. Salicylic acid (SA) is involved in seed germination, growth regulation, flower induction, thermogenesis, and biotic and abiotic stress responses in plants (Peng, Yang, Li, & Zhang, 2021). SA at all concentrations considerably improved fruit anthocyanin content (Blanch, Gómez-Jiménez, & Del Castillo, 2020; Oraei, Panahirad, Zaare-Nahandi, & Gohari, 2019). Salicylate glycosyltransferase (SGT) transfers UDP-glucose to SA and consumes SA to form a glucoside and a glucose ester (Dean & Delaney, 2008; Hu et al., 2022). SA positively regulates anthocyanin, while SGT consumes SA. It could be inferred that a reverse regulation might exist between SGT and anthocyanin. In Fig. 6 and Table S4, the strong correlations and negative R value between ArSGT1, ArSGT4, and anthocyanin derivatives seem to confirm the aforementioned inference. In our previous study, ArSGT1 was found to interact with ArMYB89 that regulates anthocyanin synthesis in red maple (Z. Chen et al., 2021), which also confirmed the regulatory effect of SGT and anthocyanins.

**Conclusion**

In summary, the present study is the first report on genome-wide characterization of GT1 family genes in red maple. A total of 560 GT genes were identified on the red maple genome. Of them, 122 members belonged to the GT1 family. All these members were unevenly distributed across 19 chromosomes, and most of them were located in the chloroplast. The number of exons in these GT1 genes ranged from 1 to 16. Most of the 122 GT1 proteins in red maple contained the GT-GTB-type domain and GT1-Gtf-like domain. Eighteen GT1 proteins in red maple might have played pivotal evolutionary roles. Network analysis indicated that the regulatory effect of GT1 family genes on anthocyanin in red maple leaves could be categorized as direct and indirect regulation. These results provide new insights into the role of GT1 family members in anthocyanin biosynthesis in red maple leaves and laid a foundation for the further functional analysis of GT1 members in Acer plants.

**Declarations**

**Ethics approval and consent to participate statement**

The authors declare that the research was carried out in accordance with a named standard.

**Funding**
Conflicts of interest
The authors declare no conflict of interest.

Compliance with Ethical Requirements
All authors Compliance with Ethical Requirements.

Availability of data and material
The datasets generated and analyzed during the current study (SRA accession number PRJNA531583) has been released on NCBI (https://www.ncbi.nlm.nih.gov/). The metabolomics and metadata reported in this paper are available at www.ebi.ac.uk/metabolights/MTBLS903 study identifier MTBLS903. Data supporting the findings of this study are available from the corresponding author (Xiaoyu Lu) upon request. The red maple genome has been deposited under BioProject Accession number: PRJNA741546.

Code availability
Not applicable.

Authors' contributions
Conceptualization: Xiaoyu Lu; Methodology: Hao Xu and Qing Zhu; Formal analysis and investigation: Hao Xu; Writing - original draft preparation: Hao Xu and Qing Zhu; Writing - review and editing: Qing Zhu and Xiaoyu Lu; Funding acquisition: Qing Zhu; Resources: Xiaoyu Lu.

Acknowledgements
This work was supported by 2019 Anhui University Humanities and Social Sciences Research Project (SK2019A1021).

References


**Figures**

**Figure 1**

Volcano maps of RL vs GL (A), RL vs YL (B), and YL vs GL (C) groups of GT genes in red maple. Red dots indicate upregulated genes, green dots indicate downregulated genes, and gray dots indicate genes with insignificant differences.
Figure 2

Chromosome location analysis of GT1 genes in red maple.

Figure 3

(A) Neighbor-joining phylogenetic tree of GT1 family proteins from red maple and *Arabidopsis*. (B) Heatmap of the expression level of GT1 genes in red maple.
Figure 4

Conserved motifs (A), domains (B), and exon–intron gene structures (C) of GT1 family proteins in red maple.

Figure 5

Collinearity analysis between the GT1 family of red maple and six representative plant species.
Figure 6

Interactive network between GT1 genes and anthocyanin derivatives in red maple.
Figure 7

Analysis of UGT78D2 in red maple. (A) Catalytic reaction involving UGT78D2. (B) Sequence alignment of UGT78D2 protein in red maple and *Arabidopsis*. 
Figure 8

Analysis of UGT73B2 in red maple. (A) Catalytic reaction involving UGT73B2. (B) Anthocyanin metabolic pathway. (C) Sequence alignment of the UGT73B2 protein in red maple and Arabidopsis.

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

- FigS1.tif
- FigS2.tif
- TableS1.xlsx
- TableS2.xlsx
- TableS3.xlsx
- TableS4.xlsx