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.

Genome-wide characterization of the GT8 gene family in Isatis indigotica and expression profiling under abiotic stress

Yifei Huang
Yili Normal University
Zhen Liu
Yili Normal University
Rong Chen
Yili Normal University
Tingting Zuo
ttzuoylsf@163.com

Yili Normal University

Research Article

Keywords: Isatis indigotica, glycosyltransferase family 8, Genome-wide analysis, abiotic stress, RT-qPCR, subcellular localization

Posted Date: November 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11400263/v1

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

Additional Declarations: No competing interests reported.

Genome-wide characterization of the GT8 gene family in *Isatis indigotica* and expression profiling under abiotic stress

Yifei Huang[1], Zhen Liu[1,2], Rong Chen[1*], Tingting Zuo[1,2*]

[1] College of Biological Sciences and Technology, Yili Normal University, Yining 835000, China

[2] Research Centre for Engineering and Technology for the Utilisation of Xinjiang's Distinctive Aromatic Plant Resources, Yining 835207, China

*Correspondence: chentianyigl@126.com and ttzuoylsf@163.com

Abstract

Background: Glycosyltransferases (GTs) are a superfamily of enzymes involved in glycosylation reactions, with the GT8 playing a key role in plant growth, development, and responses to abiotic stress. *Isatis indigotica* is a medicinal plant widely cultivated in China. It is significance to investigate how GT8 regulates abiotic stress responses in *Isatis indigotica* and to promote plant growth and development.

Results: In this study, we identified 41 GT8 genes in *Isatis indigotica* through genome-wide analysis and classified them into the GAUT, GATL, PGSIP, and GolS subfamilies. Abscisic acid-responsive elements (ABREs) and drought-associated MYB-binding sites (MBSs) were predicted in the putative promoter regions of 32 and 18 *IiGT8* genes, respectively. Functional enrichment further associated IiGT8 proteins with carbohydrate metabolism, glycosyltransferase activity, and Golgi-related functions. Analysis of public abiotic-stress transcriptome data and RT-qPCR experiments highlighted *IiGT8-03* as a candidate gene associated with salt- and drought-related conditions. IiGT8-03–GFP fluorescence was predominantly cytoplasmic in transiently transformed *Arabidopsis* mesophyll protoplasts.

Conclusions: This study characterizes the genomic organization, phylogenetic relationships, and expression patterns of the *IiGT8* gene family and identifies *IiGT8-03* as a candidate for further functional studies of abiotic stress responses.

Keywords: *Isatis indigotica*; glycosyltransferase family 8; Genome-wide analysis; abiotic stress; RT-qPCR; subcellular localization

Background

Isatis indigotica is a biennial herb of the Brassicaceae family widely used in traditional Chinese medicine. The therapeutic effects of the dried leaves and roots, known as Daqingye and Banlangen respectively, are attributed to diverse bioactive constituents, including alkaloids, flavonoids, and glycosides [1, 2]. As demand for high-quality medicinal materials grows, *Isatis indigotica* has been widely cultivated in northwestern China. However, drought and salt stresses have become major environmental constraints on its growth, yield, and quality [3–6]. Elucidating the molecular mechanisms underlying abiotic stress adaptation in *Isatis indigotica* would contribute to identifying candidate genes for improving abiotic stress resistance.

Plants adapt to abiotic stresses through physiological and metabolic adjustments, involving osmotic regulation, redox homeostasis, and cell wall remodeling [7]. Glycosyltransferases (GTs) form a large superfamily that plays a crucial role in glycan biosynthesis. Glycosyltransferase family 8 (GT8) is a major GT family in plants and contains several functionally distinct subfamilies, which are classified into galacturonosyltransferases (GAUTs), galacturonosyltransferase-like proteins (GATLs), galactinol synthases (GolSs), and plant glycogenin-like starch initiation proteins (PGSIPs) in *Arabidopsis thaliana* [8]. Genome-wide studies have characterized the GT8 gene family in *Arabidopsis thaliana* and *Populus trichocarpa* [8], *Oryza sativa* [9], *Solanum lycopersicum* [10], and *Eucalyptus*

grandis [11]. Accumulating evidence indicates that members of the GT8 family participate in plant responses to a wide range of abiotic stresses [12, 13]. Notably, *GolS* genes have attracted particular attention for their involvement in plant abiotic stress responses [14]. Raffinose family oligosaccharides (RFOs), synthesized with the participation of GolS, are important forms of carbohydrates in plants that act as compatible solutes and antioxidants, helping to protect plant cells against dehydration and oxidative damage [15–18]. GolS genes from several plant species exhibit strong stress-responsive expression [19]. In *Arabidopsis thaliana*, *AtGolS1* and *AtGolS2* are induced by drought and salinity, whereas *AtGolS3* responds mainly to cold stress [20]. Similarly, virus-induced gene silencing of *SlGolS1* increased sensitivity to cold stress in tomato, whereas overexpression of *CsGolS6* from citrus improves drought and salt tolerance in tobacco [10, 21]. *MhGolS2* has been linked to saline-alkali tolerance in apple [22]. Together, these findings support the involvement of selected GolS genes in abiotic stress tolerance and provide a rationale for examining GolS-clade members of the *Isatis indigotica* GT8 family.

Despite increasing knowledge of the molecular responses of *Isatis indigotica* to abiotic stresses, the GT8 gene family has not yet been systematically characterized in this species. In particular, the genomic organization, evolutionary relationships, and potential functional differentiation of different IiGT8 subfamilies remain unclear. Moreover, it is unknown which IiGT8 members are preferentially recruited in response to drought and salt stresses, and stress-responsive candidates inferred from transcriptomic analyses require further experimental support. Here, we identified GT8 genes in *Isatis indigotica* and analyzed their phylogenetic relationships, gene structures, conserved motifs, chromosomal distribution, synteny, and promoter cis-acting elements. Public RNA-seq datasets were used to examine expression patterns across tissues and under stress treatments. We then measured the expression of 11 selected genes by RT-qPCR during NaCl and PEG treatments and examined

the subcellular localization of *IiGT8-03*. This study aimed to characterize the genomic organization and phylogenetic relationships of the *IiGT8* family, examine expression patterns across tissues and during NaCl and PEG treatments, and identify candidate genes for further functional studies.

Methods

Plant material and abiotic stress treatments

Seeds of *Isatis indigotica* were obtained from the principal production base of the Yining Wankehui Planting Farmers' Cooperative in the Sulagong Village area, Panjin Town, Yining, Xinjiang, China (approximate central coordinates: 44.02409° N, 81.37991° E), and plant identity was authenticated by Prof. Rong Chen. Dehulled seeds were soaked in distilled water and germinated on moistened filter paper in Petri dishes at 25 °C. After radicle emergence, uniform seedlings were transferred to carbonized cotton substrate and supplied with half-strength Hoagland solution in an RXZ-280 artificial-climate chamber (Ningbo Jiangnan Instrument Factory, Ningbo, China) under a 12-h light/12-h dark photoperiod, with light/dark temperatures of 25/20 °C, a light intensity of 15,000 lx, and 60–70% relative humidity. Five-week-old seedlings at a similar developmental stage were randomly allocated to treatment groups. Salt stress was imposed by adding 150 mM NaCl to the nutrient solution, and osmotic stress was imposed with 20% (w/v) polyethylene glycol 6000 (PEG 6000). Leaves were collected at 0, 3, 6, 12, 24, and 48 h after treatment. Three seedlings were sampled independently at each time point (n=3). Samples were frozen immediately in liquid nitrogen and stored at -80 °C.

Identification and physicochemical characterization of IiGT8 proteins

Coding sequences (CDSs) were extracted from the *Isatis indigotica* genome assembly using the corresponding GTF/GFF3 annotation and the sequence-extraction module of TBtools

v2.4.68 [23, 24]. The extracted CDSs were translated into protein sequences. Candidate GT8 proteins were identified by complementary hidden Markov model (HMM) and BLASTP searches. The HMM profile of the GT8 domain (PF01501) was downloaded from the Pfam (https://www.ebi.ac.uk/interpro/entry/pfam/) and used to search the *Isatis indigotica* proteome with the TBtools Simple HMM Search function at an E-value threshold of 1×10^{-5} [25, 26]. In parallel, *Arabidopsis thaliana* GT8 reference proteins were used as BLASTP queries against the *Isatis indigotica* proteome at the same threshold. Candidate proteins were retained provided that both approaches recovered them. GT8-related domain annotations were examined in the NCBI Conserved Domain Database (CDD; https://www.ncbi.nlm.nih.gov/cdd/) [27]. The retained genes were named from *IiGT8-01* to *IiGT8-41* according to chromosomal position. Protein length, molecular weight, theoretical isoelectric point, instability index, aliphatic index, and grand average of hydropathicity were calculated with ExPASy ProtParam (https://web.expasy.org/protparam/) [28]. Subcellular localization was predicted with DeepLoc-2.1 (https://services.healthtech.dtu.dk/services/DeepLoc-2.1/) [29, 30]. Protein properties, genomic coordinates, and the best CDD hits are provided in **Additional file 1: Table S1**, **Additional file 2: Table S2**, and **Additional file 3: Table S3**, respectively.

Phylogenetic, conserved-motif, domain, and gene-structure analyses

Multiple sequence alignment of the IiGT8 and *Arabidopsis thaliana* GT8 reference proteins was used to construct a maximum-likelihood tree in IQ-TREE (LG+G4 model; 1000 bootstrap replicates), and visualized in Interactive Tree Of Life (iTOL; https://itol.embl.de/) [31, 32]. Conserved motifs in the IiGT8 protein sequences were identified using MEME through the Simple MEME Wrapper in TBtools, with the maximum number of motifs set to 10 and all remaining parameters set to their default values [33]. Conserved domains identified

by the NCBI Conserved Domain Database were incorporated into composite visualization. Exon–intron structures were visualized using the TBtools Gene Structure View function.

Promoter cis-regulatory element analysis

The 2000-bp genomic sequence upstream of the start codon of each IiGT8 gene was extracted with TBtools and treated as the putative promoter. Putative cis-regulatory elements were identified in these upstream sequences using PlantCARE [34]. Predicted elements were grouped according to their annotated functions, and their counts were summarized and visualized in R v4.4.1. The element-count matrix is provided in **Additional file 4: Table S4**.

Chromosomal distribution, gene duplication, and synteny analyses

Gene density and physical positions of IiGT8 genes were obtained from the *Isatis indigotica* annotation with the TBtools Gene Density Profile and Gene Location Visualize (GTF/GFF) modules. Intraspecific collinear blocks in *Isatis indigotica* and interspecific collinear blocks between *Isatis indigotica* and *Arabidopsis thaliana* were identified using MCScanX in TBtools. MCScanX was run with MATCH_SCORE = 50, MATCH_SIZE = 5, GAP_PENALTY = −1, OVERLAP_WINDOW = 5, MAX_GAPS = 25, and an alignment E-value threshold of 1×10^{-5}. Intraspecific relationships were visualized using Advanced Circos [35]. Tandem duplicates were defined as adjacent paralogs separated by ≤ 200 kb. Interspecific syntenic relationships between *Isatis indigotica* and *Arabidopsis thaliana* were visualized using the Dual Synteny Plot module of TBtools. Nonsynonymous (Ka) and synonymous (Ks) substitution rates and Ka/Ks ratios of collinear *IiGT8* gene pairs were estimated from their coding sequences using the TBtools Simple Ka/Ks Calculator [36]. Substitution-rate estimates are supplied in **Additional file 5: Table S5**.

Analysis of public RNA-seq datasets

Raw RNA-seq reads were assessed using FastQC and processed with Trimmomatic for adapter removal and quality filtering [37]. Transcript abundance was estimated using kallisto and expressed as transcripts per million (TPM) [38]. *IiGT8* expression profiles were retrieved by matching gene identifiers. Genes with low or no expression were excluded, and the remaining TPM values were standardized to row-wise Z scores and clustered for heatmap visualization in R v4.4.1. Candidate genes were selected to represent multiple phylogenetic groups and contrasting transcriptome response patterns for subsequent RT-qPCR analysis.

Functional enrichment and predicted protein-association network

Potential functional associations among IiGT8 proteins were analyzed with STRING (https://string-db.org/) with a minimum interaction score of 0.150 [39]. GO and KEGG enrichment analyses were performed. Terms with a false discovery rate < 0.05 were considered enriched. The STRING network was imported into Cytoscape (v3.8.0) for visualization. Edges represent predicted functional associations rather than evidence of direct physical interaction. Full enrichment output is provided in **Additional file 6: Table S6**.

RNA extraction and RT-qPCR validation

Total RNA was extracted from *Isatis indigotica* leaves using the Polysaccharides & Polyphenolics-rich Plant Total RNA Kit (SimGen, Hangzhou, China) according to the manufacturer's instructions. RNA integrity was assessed by electrophoresis on a 1.5% (w/v) agarose gel, and RNA purity was evaluated using a K5600C microvolume spectrophotometer (KAIAO, Beijing, China). RNA samples with A_{260}/A_{280} ratios of 1.9–2.1 and A_{260}/A_{230} ratios of 2.0–2.3 were used for reverse transcription. First-strand cDNA was synthesized using the All-in-One First-Strand cDNA Synthesis Kit (gDNA Purge) (SimGen, Hangzhou, China) and subsequently diluted tenfold with nuclease-free water. RT-qPCR was performed on a CFX96

Real-Time PCR Detection System (Bio-Rad, Hercules, CA, USA) with SYBR Green PCR Mix (SimGen, Hangzhou, China). *IiMUB* and *IiACT* were used as reference genes, and their geometric mean was used for normalization and Gene-specific primers were designed using Primer Premier 5.0. Primer sequences are provided in **Additional file 7: Table S7**. Each group comprised three biological replicates, each analyzed in two technical replicates. Technical replicates were averaged before statistical analysis. Relative expression was calculated with the $2^{-\Delta\Delta Ct}$ method using the 0-h group as the calibrator.

Construction of the *IiGT8-03*–GFP vector and transient expression

The *IiGT8-03*-GFP fusion and the empty-vector control were expressed in mesophyll protoplasts from 4-week-old *Arabidopsis thaliana* leaves by PEG-Ca^{2+}-mediated transformation as previously described [40]. mCherry-SYP131$^{\Delta TM}$ served as a cytoplasmic marker where indicated[41]. Protoplasts were incubated in darkness at 23 °C for 16 h and examined with an Olympus FV3000 laser-scanning confocal microscope. Localization was assessed by comparing the distribution of *IiGT8-03*–GFP with those of the empty vector control, chloroplast autofluorescence, and red fluorescent marker.

Statistical analysis

All data are presented as the mean ± standard error of the mean (SEM). For each gene, relative expression values ($2^{-\Delta\Delta Ct}$) were analyzed separately for the NaCl and PEG time courses using one-way ANOVA followed by Dunnett's multiple-comparisons test, with each post-treatment time point compared with the 0-h group. These comparisons assessed changes relative to the pretreatment baseline; time-matched untreated controls were not included. Statistical significance is indicated as $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, and $****p < 0.0001$. Analyses were performed in GraphPad Prism v10.1.2.

Results

Identification and physicochemical characteristics of the IiGT8 family

Genome-wide screening identified 41 candidate GT8 genes in *Isatis indigotica*, designated *IiGT8-01* to *IiGT8-41* according to their genomic positions (**Additional file 1: Table S1; Additional file 2: Table S2**). The predicted proteins ranged from 105 amino acids for IiGT8-29 to 1056 amino acids for IiGT8-08, with molecular weights of 12.88–118.66 kDa and theoretical isoelectric points of 4.81–9.66. Fifteen proteins had pI values below 7, whereas 26 had values above 7. Negative GRAVY values were obtained for 38 proteins. DeepLoc-2.1 assigned 37 proteins to a single compartment: 24 to the Golgi apparatus, seven to the cytoplasm, three to the endoplasmic reticulum and three to the extracellular space. The remaining four proteins had dual localization predictions. IiGT8-03 and IiGT8-25 were among the proteins predicted to localize to the cytoplasm (**Additional file 1: Table S1**).

Phylogenetic relationships of IiGT8 proteins

The maximum-likelihood tree separated the 41 IiGT8 proteins into four subfamilies: GAUT, GATL, PGSIP and GolS, containing 17, seven, eight and nine members, respectively (**Fig. 1**). GAUT was the largest group, accounting for 41.5% of the family. IiGT8 proteins clustered with Arabidopsis reference proteins within each subfamily. In the GolS branch, IiGT8-13 grouped with AtGolS3 and IiGT8-25 with AtGolS1, with bootstrap support of 100% and 97%, respectively. *IiGT8-03* grouped with AtGolS2, although support for this terminal pairing was lower, at 58%.

Conserved motifs, domain architecture and gene structure

The ten motifs identified by MEME formed distinct combinations across the IiGT8 proteins (**Fig. 2A**). IiGT8-03 and IiGT8-13 shared the motif order 7–1–9–6–8, whereas IiGT8-07 and IiGT8-23 shared the order 10–2–6–1–4–5–3. A Glyco_transf_8 (PF01501) match was

detected in all 41 candidates, with variation in the position and extent of the matched region (**Fig. 2B; Additional file 3: Table S3**). The shortest candidate, IiGT8-29, had a PF01501 match spanning residues 9–63 of its 105-amino-acid sequence.

Exon–intron organization also differed among members (**Fig. 2C**). *IiGT8-03* and *IiGT8-13* each had four CDS-containing exons, but the first intron was longer in *IiGT8-03*. *IiGT8-07* and *IiGT8-23* each had six CDS-containing exons and similar exon arrangements.

Cis-regulatory elements in putative IiGT8 promoter regions

The 2000-bp sequences upstream of the start codons of the 41 *IiGT8* genes contained predicted elements associated with light responses, growth and development, hormone responses and environmental stress (**Fig. 3; Additional file 4: Table S4**). ABREs were predicted in the putative promoter regions of 32 of the 41 genes (78.0%), and drought-associated MYB-binding sites (MBSs) were predicted in those of 18 genes (43.9%). Methyl jasmonate-associated CGTCA and/or TGACG motifs were detected in 25 genes, while low-temperature-responsive elements (LTRs) occurred in 23 genes. Element counts also varied among individual genes. The putative promoter regions of *IiGT8-03* and *IiGT8-25* contained eight and seven predicted ABREs, respectively. The putative promoter region of *IiGT8-01* contained 14 LTRs, the highest LTR count in the family.

Chromosomal distribution and synteny of IiGT8 genes

Forty *IiGT8* genes were located on the seven chromosomes, whereas *IiGT8-41* was located on scaffold Contig00526 (**Fig. 4A; Additional file 2: Table S2**). Chromosome 3 carried the most members (eight), and chromosome 2 carried the fewest (four). On chromosome 1, *IiGT8-04*, *IiGT8-05* and *IiGT8-06* occupied a 20.7-kb interval. Interspecific analysis identified 39 collinear gene pairs involving 28 *IiGT8* genes and 30 Arabidopsis GT8 genes (**Fig. 4B; Additional file 8: Table S8**). Three segmentally duplicated pairs were identified

within the *Isatis indigotica* genome: *IiGT8-03*/*IiGT8-13*, *IiGT8-01*/*IiGT8-18* and *IiGT8-07*/*IiGT8-23* (**Fig. 4C**). Their Ka/Ks ratios were 0.119, 0.190 and 0.056, respectively, and were below 1 in all three cases (**Additional file 5: Table S5**).

IiGT8 expression profiles across treatments and tissues

Public RNA-seq profiles, expressed as gene-wise Z-scores, showed differences in *IiGT8* expression across treatments and tissues (**Fig. 5**). In the dataset comparing control, drought and salt conditions, *IiGT8-11* showed higher relative expression in all three salt-treated samples than in the controls (**Fig. 5A**). In the separate salt time course, *IiGT8-03* showed higher relative expression at 48 h than in the control samples (**Fig. 5B**). In the tissue dataset, *IiGT8-03* expression was higher in fruit than in roots or leaves, whereas *IiGT8-26* expression was higher in leaves than in the other tissues examined (**Fig. 5C**).

Eleven genes were selected for RT-qPCR to represent all four subfamilies and a range of transcriptome expression patterns. The set comprised six GAUT members (*IiGT8-07*, *IiGT8-14*, *IiGT8-15*, *IiGT8-19*, *IiGT8-38* and *IiGT8-39*), two GATL members (*IiGT8-01* and *IiGT8-10*), one PGSIP member (*IiGT8-17*) and two GolS members (*IiGT8-03* and *IiGT8-25*).

Functional enrichment and predicted protein associations

GO enrichment linked the IiGT8 set to several aspects of carbohydrate metabolism (**Fig. 6A; Additional file 6: Table S6**). Among biological-process terms, polysaccharide biosynthesis involved 30 genes (FDR = 3.53×10^{-48}), with additional enrichment in pectin metabolism, cell-wall organization and xylan biosynthesis. Molecular-function terms included UDP-glycosyltransferase activity and inositol 3-α-galactosyltransferase activity; the latter involved the nine GolS-clade members (FDR = 1.28×10^{-19}). Golgi membrane was a prominent cellular-component term, involving 28 genes. KEGG analysis identified enrichment in amino

sugar and nucleotide sugar metabolism (24 genes), galactose metabolism (nine genes; FDR = 3.04×10^{-12}) and the broader metabolic pathways category (**Fig. 6B**).

The predicted association network contained a densely connected group of IiGT8 proteins at the STRING score threshold of 0.150 (**Fig. 7**). IiGT8-15, IiGT8-01 and IiGT8-26 showed greater connectivity than the GolS-clade members IiGT8-03 and IiGT8-25, which occupied peripheral positions. These edges represent predicted functional associations rather than experimentally confirmed physical interactions.

Temporal expression profiles of selected IiGT8 genes

RT-qPCR revealed different temporal profiles among the 11 selected genes during treatment with 150 mM NaCl or 20% PEG-6000 (**Fig. 8**). Expression at each sampling time was normalized to the corresponding 0-h baseline. During NaCl treatment, *IiGT8-03* reached approximately 4.8-fold at 6 h and 8.5-fold at 24 h, before falling below baseline at 48 h. *IiGT8-10* reached a transient peak of approximately 6.4-fold at 6 h; *IiGT8-07* and *IiGT8-14* also reached their highest observed values at 6 h, at approximately two- to threefold. *IiGT8-39* reached approximately 3.8-fold at 6 h. In contrast, *IiGT8-01* and *IiGT8-25* remained below baseline throughout the post-treatment sampling period.

During PEG treatment, *IiGT8-03* showed an early rise followed by a decline and a second increase. Relative expression reached approximately 57.7-fold at 6 h, decreased to 3.4-fold at 24 h, and then rose to 40.3-fold at 48 h. *IiGT8-10*, *IiGT8-07* and *IiGT8-17* showed transient peaks at 6 h of approximately 24-, 17- and 10-fold, respectively. *IiGT8-39* remained at approximately fourfold from 3 to 12 h. *IiGT8-25* followed a contrasting course, reaching approximately 1.6-fold at 3 h before dropping below baseline from 6 to 48 h.

The two GolS-clade candidates, *IiGT8-03* and *IiGT8-25*, differed in the timing and magnitude of their expression changes relative to their respective 0-h baselines (**Fig. 8**). The pronounced

changes in *IiGT8-03* during both time courses supported its selection for subcellular-localization analysis. Because time-matched untreated controls were not included, these profiles describe changes from 0 h rather than treatment effects isolated from temporal variation.

Subcellular localization of IiGT8-03

IiGT8-03–GFP fluorescence was concentrated mainly in the peripheral cytoplasmic layer of Arabidopsis mesophyll protoplasts and broadly overlapped with the mCherry–SYP131$^{\Delta TM}$ cytoplasmic marker (**Fig. 9**). Its distribution was distinct from chloroplast autofluorescence. The empty-vector GFP control also produced diffuse cytoplasmic and peripheral fluorescence, while the *IiGT8-03* fusion showed no discrete punctate or organelle-specific enrichment. These observations were consistent with the predicted cytoplasmic localization of *IiGT8-03* under the conditions tested.

Discussion

The IiGT8 family combines conserved subfamily architecture with diverse transcriptional profiles across tissues and treatment time courses. Integrating these features with localization evidence identifies *IiGT8-03* as a priority candidate for studying carbohydrate metabolism during abiotic stress. Its contrasting expression trajectory relative to *IiGT8-25* points to transcriptional differentiation within the GolS clade. These findings connect family evolution with the selection of individual genes for biochemical and genetic analysis.

The GAUT, GATL, PGSIP and GolS divisions in *Isatis indigotica* correspond to the major plant GT8 lineages described by Yin et al. [8]. Similar organization in grasses, tomato and *Eucalyptus grandis* indicates conservation of these lineages across plant species [9–11]. The 41 identified members are comparable in number to the 40 reported in tomato and fewer than the 52 in *E. grandis* [10, 11]. Within *Isatis indigotica*, concordant phylogenetic groups, motif

arrangements and exon–intron structures further support structural conservation within subfamilies.

Ka/Ks ratios below 1 in all three segmentally duplicated pairs are consistent with purifying selection acting on retained IiGT8 duplicates. Gene duplication and differential retention have also shaped GT8 evolution in grasses [9]. Such coding-sequence conservation can coexist with divergence in transcriptional regulation or catalytic properties. The *IiGT8-03*/*IiGT8-13* pair offers a specific context for examining this relationship within a duplicated GolS lineage.

The enrichment results reflect the biochemical diversity of the GT8 family. Some *Arabidopsis* GAUT proteins synthesize homogalacturonan [42], whereas GUX1, GUX2 and GUX4, historically assigned to the PGSIP group, exhibit xylan glucuronosyltransferase activity [43]. Loss of *Arabidopsis GAUT10* alters root cell-wall polysaccharide composition and reduces root growth under sucrose-limited conditions [44]. These activities connect individual GT8 members to distinct aspects of cell-wall metabolism and growth. This diversity helps explain the concurrent enrichment of carbohydrate metabolism, cell wall-related processes and Golgi-associated terms in the IiGT8 family. Predicted functional associations and temporal expression profiles provide complementary criteria for selecting members of these different subfamilies for further study.

Promoter composition provides a basis for exploring transcriptional differentiation within the IiGT8 family. Predicted ABRE and MBS elements suggest links to hormonal and environmental signaling. *IiGT8-03* and *IiGT8-25* carry several predicted ABREs yet follow contrasting expression trajectories, directing attention to motif arrangement, transcription factor availability and chromatin context. Regulatory studies in other species illustrate how such differences can be investigated. HY5 binds light-responsive elements in the *AtGolS1*

promoter [45], while VvDREB2A binds and activates the *VvGolS3* promoter in grapevine [46]. Testing specific transcription factor–promoter interactions would clarify the regulatory basis of the contrasting IiGT8 profiles.

The expression time courses highlight temporal differentiation between the two GolS-clade candidates. Relative to its 0-h baseline, *IiGT8-03* showed an early increase and a second rise during PEG treatment, with a later maximum during NaCl treatment. Expression of *IiGT8-25* remained below baseline at most sampling times, apart from a modest early increase during PEG treatment. These contrasting within-gene profiles provide a basis for comparing the temporal regulation of GolS-clade members. Distinct expression patterns also occur in *Arabidopsis*, where *AtGolS1* and *AtGolS2* respond to drought and salinity, whereas *AtGolS3* responds predominantly to low temperature [20]. Sampling across the treatment period therefore captures features that a single time point could miss.

The GolS-clade placement of *IiGT8-03* links its expression profile to a specific hypothesis involving galactinol metabolism. GolS catalyzes galactinol formation from UDP-galactose and myo-inositol, providing the galactosyl donor for raffinose synthesis. In *Arabidopsis*, *AtGolS2* overexpression increases galactinol and raffinose accumulation and improves drought tolerance [20]. Nishizawa et al. associated these sugars with reduced oxidative damage and demonstrated their hydroxyl radical-scavenging capacity in vitro [47]. Together, this biochemical context and the *IiGT8-03* expression profile motivate testing its proposed participation in sugar-mediated osmotic or redox regulation.

The physiological consequences of GolS activity depend on the surrounding metabolic pathway. Tobacco plants overexpressing citrus *CsGolS6* showed improved drought and salt tolerance without a significant increase in galactinol content. Some lines had lower raffinose levels alongside increases in metabolites with antioxidant properties [21]. In *Arabidopsis rs5*

357 mutants, GolS activity increased during cold treatment, but raffinose did not accumulate in
358 leaves, illustrating the importance of downstream biosynthetic steps [48]. Transgenic studies
359 have also linked apple *MhGolS2* to saline–alkali tolerance [22]. These findings make
360 metabolite profiling an essential complement to expression analysis when investigating the
361 contribution of *IiGT8-03* to stress physiology.

362 GolS activity also connects carbohydrate metabolism with carbon allocation and growth. In
363 cucumber, cold-inducible *CsGolS1* overexpression in minor veins increased raffinose family
364 oligosaccharides in phloem sap and promoted sink growth under cold stress [49]. In tomato,
365 CRISPR/Cas9 disruption of *SlGolS4* produced a semi-dwarf phenotype with lower galactinol,
366 raffinose, lignin and flavonoid levels [50]. These findings provide a rationale for examining
367 growth and carbon allocation alongside stress-related traits in future *IiGT8-03* studies.

368 The predominantly cytoplasmic distribution of IiGT8-03–GFP in *Arabidopsis* mesophyll
369 protoplasts provides a cellular context for the proposed metabolic role. The fluorescence
370 broadly overlapped with the cytoplasmic marker under transient-expression conditions.
371 Chloroplast localization reported for tomato SlGolS1 [10] illustrates variation in the cellular
372 distribution of GolS family members. Comparing these distributions with biochemical
373 activity would help connect cellular location to functional differentiation within the family.

374 Biochemical and genetic analyses can now connect candidate selection to the function of
375 *IiGT8-03*. Assays with purified protein, combined with galactinol and raffinose
376 measurements, would test the proposed GolS activity. Genetic perturbation coupled with
377 metabolite profiling and physiological measurements under salinity and soil water deficit
378 would determine its contribution to stress tolerance. Reported links between environmental
379 stress and medicinal constituents in *Isatis indigotica* [4] also motivate subsequent tests of

whether *IiGT8-03* affects medicinal quality. Future studies should test whether *IiGT8-03* contributes to stress tolerance by altering galactinol and raffinose metabolism.

Conclusions

A total of 41 GT8 genes were identified in *Isatis indigotica* and classified into four subfamilies with distinct structural features. Three segmentally duplicated gene pairs showed Ka/Ks ratios consistent with purifying selection. Expression profiling revealed variation among IiGT8 genes across tissues and treatments. In the RT-qPCR time course, *IiGT8-03* showed large increases in expression compared with the expression level at 0 h during both NaCl and PEG treatments, whereas IiGT8-25, another GolS-clade member, showed a contrasting temporal profile. Without time-matched untreated controls, these changes cannot be attributed specifically to the treatments. *IiGT8-03*–GFP displayed a predominantly cytoplasmic fluorescence pattern. These results identify *IiGT8-03* as a candidate for further investigation; its biochemical activity and contribution to stress tolerance remain to be tested.

List of abbreviations

ANOVA, analysis of variance

BLASTP, protein Basic Local Alignment Search Tool

CDD, Conserved Domain Database

FDR, false discovery rate

GATL, galacturonosyltransferase-like

GAUT, galacturonosyltransferase

GFP, green fluorescent protein

GO, Gene Ontology

GolS, galactinol synthase

GRAVY, grand average of hydropathicity

GT8, glycosyltransferase family 8

HMM, hidden Markov model

Ka, nonsynonymous substitution rate

KEGG, Kyoto Encyclopedia of Genes and Genomes

Ks, synonymous substitution rate

PEG, polyethylene glycol

PGSIP, plant glycogenin-like starch initiation protein

RFO, raffinose family oligosaccharide

RT-qPCR, reverse-transcription quantitative polymerase chain reaction

SEM, standard error of the mean

SRA, Sequence Read Archive

TAIR, The Arabidopsis Information Resource

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

The *Isatis indigotica* genome resources used in this study are available in Figshare (https://doi.org/10.6084/m9.figshare.8940377.v1 and https://doi.org/10.6084/m9.figshare.8940383.v1) and NCBI (accession VHIU00000000) [51].

The *Arabidopsis thaliana* genome assembly and annotation used for comparative analyses were obtained from The Arabidopsis Information Resource (TAIR; https://www.arabidopsis.org/) [52]. The public RNA-seq datasets are associated with BioProjects PRJNA854335 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA854335) [3] and PRJNA1400285 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1400285) [53] in NCBI, and PRJCA032613 (https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA032613) [4] and PRJCA006185 (https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA006185) [54] in the Genome Sequence Archive. Supplementary annotation and analysis tables are provided in Additional files 1–8. Other data supporting the findings of this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Third Batch of the Tianchi Innovation Leading Talent Fund of the Xinjiang Uygur Autonomous Region (2025CXLJ004); the Rural Science and Technology Entrepreneurship Action Project for Science and Technology Commissioners in the Xinjiang Uygur Autonomous Region (2025KZ014); and the Xinjiang Uygur Autonomous Region “Tianchi Yingcai” Project (2023TCYCQNBS02).

Authors’ contributions

YH, ZL and TZ conceptualized the study and performed the formal analysis. YH, ZL, RC, and TZ developed the methodology. YH, ZL and RC conducted the investigation. YH performed validation and data curation and wrote the original draft. RC provided software support. RC and TZ prepared the visualizations, acquired funding, and reviewed and edited

the manuscript. TZ supervised the study and managed the project. All authors read and approved the final manuscript.

Acknowledgements

The authors thank the researchers who generated and publicly shared the genomic and transcriptomic datasets used in this study, as well as the teams maintaining Figshare, the National Center for Biotechnology Information (NCBI), the Sequence Read Archive (SRA), The Arabidopsis Information Resource (TAIR), and the Genome Sequence Archive (GSA), for making these resources available. We also acknowledge the developers and curators of Pfam, the NCBI Conserved Domain Database (CDD), PlantCARE, STRING, Gene Ontology (GO), and the Kyoto Encyclopedia of Genes and Genomes (KEGG) for the annotation and functional resources used in our analyses.

We are grateful to the developers of TBtools and its integrated analysis and visualization modules, BLASTP, ExPASy ProtParam, DeepLoc-2.1, IQ-TREE, Interactive Tree Of Life (iTOL), MEME, MCScanX, FastQC, Trimmomatic, kallisto, R, and Cytoscape for the computational tools used in this study. We also acknowledge the developers of Primer Premier and GraphPad Prism for the software used for primer design and statistical analysis, respectively.

Additional files

Additional file 1 (.xlsx). Table S1. Identifiers, physicochemical properties, and predicted subcellular localization of IiGT8 proteins.

Additional file 2 (.xlsx). Table S2. Genomic coordinates of IiGT8 genes.

Additional file 3 (.xlsx). Table S3. Best CDD domain hit for each IiGT8 candidate.

Additional file 4 (.xlsx). Table S4. Cis-acting element count matrix for 41 IiGT8 promoters.

Additional file 5 (.xlsx). Table S5. Ka, Ks, and Ka/Ks estimates for three duplicated IiGT8 gene pairs.

Additional file 6 (.xlsx). Table S6. GO and KEGG enrichment output, including false discovery rates and matching genes.

Additional file 7 (.xlsx). Table S7. RT-qPCR primer sequences for 11 target genes and two reference genes.

Additional file 8 (.xlsx). Table S8. Intraspecific and interspecific GT8 collinear gene pairs.

References

1. Chen Q, Lan H-Y, Peng W, Rahman K, Liu Q-C, Luan X, et al. *Isatis indigotica* : a review of phytochemistry, pharmacological activities and clinical applications. Journal of Pharmacy and Pharmacology. 2021;73:1137–50. https://doi.org/10.1093/jpp/rgab014.

2. Feng J, Huang D, Yang Y, Chen J, Qiu S, Lv Z, et al. Isatis indigotica: from (ethno) botany, biochemistry to synthetic biology. Mol Horticulture. 2021;1:17. https://doi.org/10.1186/s43897-021-00021-w.

3. Jiang M, Wang Z, Ren W, Yan S, Xing N, Zhang Z, et al. Identification of the bZIP gene family and regulation of metabolites under salt stress in isatis indigotica. Front Plant Sci. 2022;13:1011616. https://doi.org/10.3389/fpls.2022.1011616.

4. Wang Z, Cao B, Du J, Deng T, Wang R, Chen Y, et al. Transcriptomic and proteomic-based analysis of the mechanisms by which drought and salt stresses affect the quality of Isatidis Folium. BMC Plant Biol. 2025;25:332. https://doi.org/10.1186/s12870-025-06309-z.

5. Wu S, Jia X, Tian B, Zhang F, Zhao J, Xie X, et al. Physiological and Cellular Ultrastructural Responses of Isatis indigotica Fort. under Salt Stress. Plants. 2024;13:1593. https://doi.org/10.3390/plants13121593.

494 6. Li W, Yao J, He C, Ren Y, Zhao L, He X. The synergy of dark septate endophytes and
495 organic residue on Isatis indigotica growth and active ingredients accumulation under
496 drought stress. Industrial Crops and Products. 2023;203:117147.
497 https://doi.org/10.1016/j.indcrop.2023.117147.
498 7. Zhou C, Yu S, Zhang H, Li F. Physiological and biochemical responses of Isatis indigotica
499 to deficit irrigation in a cold and arid environment. Frontiers in Plant Science.
500 2023;13:1094158. https://doi.org/10.3389/fpls.2022.1094158.
501 8. Yin Y, Chen H, Hahn MG, Mohnen D, Xu Y. Evolution and Function of the Plant Cell Wall
502 Synthesis-Related Glycosyltransferase Family 8. Plant Physiology. 2010;153:1729–46.
503 https://doi.org/10.1104/pp.110.154229.
504 9. Kong W, Gong Z, Zhong H, Zhang Y, Zhao G, Gautam M, et al. Expansion and
505 Evolutionary Patterns of Glycosyltransferase Family 8 in Gramineae Crop Genomes and
506 Their Expression under Salt and Cold Stresses in Oryza sativa ssp. japonica. Biomolecules.
507 2019;9:188. https://doi.org/10.3390/biom9050188.
508 10. Zhang H, Mo F, Li D, Zheng J, Liang S, Liu S, et al. Genome-wide identification and
509 expression analysis of the GT8 gene family in tomato(Solanum lycopersicum) and the
510 functional of SlGolS1 under cold stress. Scientia Horticulturae. 2024;338:113686.
511 https://doi.org/10.1016/j.scienta.2024.113686.
512 11. Jie Y, Liu L, Li L, Wu A-M, Fan C, Liu S. Comprehensive genome-wide identification
513 and functional analysis of the GT8 gene family in Eucalyptus Grandis. Front Plant Sci.
514 2025;16:1610059. https://doi.org/10.3389/fpls.2025.1610059.
515 12. Shikakura Y, Oguchi T, Yu X, Ohtani M, Demura T, Kikuchi A, et al. Transgenic poplar
516 trees overexpressing AtGolS2, a stress-responsive galactinol synthase gene derived from
517 Arabidopsis thaliana, improved drought tolerance in a confined field. Transgenic Research.

518 2022;31:579–91. https://doi.org/10.1007/s11248-022-00321-x.

519 13. Xu G, He M, Yan S, Lyu D, Cheng C, Zhao D, et al. Galactinol synthase gene 5
520 (MdGolS5) enhances the cold resistance of apples by promoting raffinose family
521 oligosaccharide accumulation. Plant Physiology and Biochemistry. 2025;220:109416.
522 https://doi.org/10.1016/j.plaphy.2024.109416.

523 14. Xiao Y, Dong Y, Zhang Y, Zhang Y, Liu L, Liu P, et al. Two galactinol synthases
524 contribute to the drought response of Camellia sinensis. Planta. 2023;258:84.
525 https://doi.org/10.1007/s00425-023-04238-5.

526 15. Dos Santos TB, Vieira LGE. Involvement of the galactinol synthase gene in abiotic and
527 biotic stress responses: A review on current knowledge. Plant Gene. 2020;24:100258.
528 https://doi.org/10.1016/j.plgene.2020.100258.

529 16. Salvi P, Kamble NU, Majee M. Stress-Inducible Galactinol Synthase of Chickpea
530 (CaGolS) is Implicated in Heat and Oxidative Stress Tolerance Through Reducing Stress-
531 Induced Excessive Reactive Oxygen Species Accumulation. Plant and Cell Physiology.
532 2018;59:155–66. https://doi.org/10.1093/pcp/pcx170.

533 17. ElSayed AI, Rafudeen MS, Golldack D. Physiological aspects of raffinose family
534 oligosaccharides in plants: protection against abiotic stress. Plant Biol J. 2014;16:1–8.
535 https://doi.org/10.1111/plb.12053.

536 18. Sengupta S, Mukherjee S, Basak P, Majumder AL. Significance of galactinol and
537 raffinose family oligosaccharide synthesis in plants. Front Plant Sci. 2015;6:656.
538 https://doi.org/10.3389/fpls.2015.00656.

539 19. de Koning R, Wils GE, Kiekens R, De Vuyst L, Angenon G. Impact of drought and salt
540 stress on galactinol and raffinose family oligosaccharides in common bean (*Phaseolus*
541 *vulgaris*). AoB PLANTS. 2023;15:plad038. https://doi.org/10.1093/aobpla/plad038.

542 20. Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, et al. Important roles of
543 drought- and cold-inducible genes for galactinol synthase in stress tolerance in *Arabidopsis*
544 *thaliana*. The Plant Journal. 2002;29:417–26. https://doi.org/10.1046/j.0960-
545 7412.2001.01227.x.

546 21. Martins CPS, Fernandes D, Guimarães VM, Du D, Silva DC, Almeida A-AF, et al.
547 Comprehensive analysis of the GALACTINOL SYNTHASE (GolS) gene family in citrus and
548 the function of CsGolS6 in stress tolerance. PLoS ONE. 2022;17:e0274791.
549 https://doi.org/10.1371/journal.pone.0274791.

550 22. Wang X, Zhang Z, Li J, Wang Y. Genome-wide analysis of the GT8 gene family in apple
551 and functional identification of MhGolS2 in saline-alkali tolerance. Plant Mol Biol.
552 2024;114:103. https://doi.org/10.1007/s11103-024-01499-w.

553 23. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: An Integrative
554 Toolkit Developed for Interactive Analyses of Big Biological Data. Molecular Plant.
555 2020;13:1194–202. https://doi.org/10.1016/j.molp.2020.06.009.

556 24. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, et al. TBtools-II: A “one for all, all for one”
557 bioinformatics platform for biological big-data mining. Molecular Plant. 2023;16:1733–42.
558 https://doi.org/10.1016/j.molp.2023.09.010.

559 25. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al.
560 Pfam: The protein families database in 2021. Nucleic Acids Research. 2021;49:D412–9.
561 https://doi.org/10.1093/nar/gkaa913.

562 26. Paysan-Lafosse T, Andreeva A, Blum M, Chuguransky SR, Grego T, Pinto BL, et al. The
563 Pfam protein families database: embracing AI/ML. Nucleic Acids Research. 2025;53:D523–
564 34. https://doi.org/10.1093/nar/gkae997.

565 27. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved

566 domain database in 2023. Nucleic Acids Research. 2023;51:D384–8.
567 https://doi.org/10.1093/nar/gkac1096.
568 28. Duvaud S, Gabella C, Lisacek F, Stockinger H, Ioannidis V, Durinx C. Expasy, the Swiss
569 Bioinformatics Resource Portal, as designed by its users. Nucleic Acids Research.
570 2021;49:W216–27. https://doi.org/10.1093/nar/gkab225.
571 29. Nielsen H. Practical Applications of Language Models in Protein Sorting Prediction:
572 SignalP 6.0, DeepLoc 2.1, and DeepLocPro 1.0. In: Kc DB, editor. Large Language Models
573 (LLMs) in Protein Bioinformatics. New York, NY: Springer US; 2025. p. 153–75.
574 https://doi.org/10.1007/978-1-0716-4623-6_10.
575 30. Ødum MT, Teufel F, Thumuluri V, Almagro Armenteros JJ, Johansen AR, Winther O, et
576 al. DeepLoc 2.1: multi-label membrane protein type prediction using protein language
577 models. Nucleic Acids Research. 2024;52:W215–20. https://doi.org/10.1093/nar/gkae237.
578 31. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, et
579 al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the
580 Genomic Era. Molecular Biology and Evolution. 2020;37:1530–4.
581 https://doi.org/10.1093/molbev/msaa015.
582 32. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic
583 tree display and annotation tool. Nucleic Acids Research. 2024;52:W78–82.
584 https://doi.org/10.1093/nar/gkae268.
585 33. Bailey TL, Johnson J, Grant CE, Noble WS. The MEME Suite. Nucleic Acids Res.
586 2015;43:W39–49. https://doi.org/10.1093/nar/gkv416.
587 34. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. PlantCARE, a
588 database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of
589 promoter sequences. Nucleic Acids Research. 2002;30:325–7.

https://doi.org/10.1093/nar/30.1.325.

35. Wang Y, Tang H, Wang X, Sun Y, Joseph PV, Paterson AH. Detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. Nat Protoc. 2024;19:2206–29. https://doi.org/10.1038/s41596-024-00968-2.

36. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics, Proteomics & Bioinformatics. 2010;8:77–80. https://doi.org/10.1016/S1672-0229(10)60008-3.

37. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.

38. Sullivan DK, Min KH, Hjörleifsson KE, Luebbert L, Holley G, Moses L, et al. kallisto, bustools and kb-python for quantifying bulk, single-cell and single-nucleus RNA-seq. Nat Protoc. 2025;20:587–607. https://doi.org/10.1038/s41596-024-01057-0.

39. Szklarczyk D, Nastou K, Koutrouli M, Kirsch R, Mehryary F, Hachilif R, et al. The STRING database in 2025: protein networks with directionality of regulation. Nucleic Acids Research. 2025;53:D730–7. https://doi.org/10.1093/nar/gkae1113.

40. Yoo S-D, Cho Y-H, Sheen J. Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nature Protocols. 2007;2:1565–72. https://doi.org/10.1038/nprot.2007.199.

41. Li B, Li Y, Liu F, Tan X, Rui Q, Tong Y, et al. Overexpressed Tomosyn Binds Syntaxins and Blocks Secretion during Pollen Development. Plant Physiol. 2019;181:1114–26. https://doi.org/10.1104/pp.19.00965.

42. Engle KA, Amos RA, Yang J, Glushka J, Atmodjo MA, Tan L, et al. Multiple Arabidopsis galacturonosyltransferases synthesize polymeric homogalacturonan by oligosaccharide acceptor-dependent or *de novo* synthesis. The Plant Journal. 2022;109:1441–56.

https://doi.org/10.1111/tpj.15640.

43. Rennie EA, Hansen SF, Baidoo EEK, Hadi MZ, Keasling JD, Scheller HV. Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. Plant Physiology. 2012;159:1408–17. https://doi.org/10.1104/pp.112.200964.

44. Dash L, Swaminathan S, Šimura J, Gonzales CLP, Montes C, Solanki N, et al. Changes in cell wall composition due to a pectin biosynthesis enzyme GAUT10 impact root growth. Plant Physiology. 2023;193:2480–97. https://doi.org/10.1093/plphys/kiad465.

45. Ranjan A, Michael R, Gautam S, Trivedi PK. HY5-dependent light-mediated regulation of galactinol synthase gene, AtGolS1, modulates galactinol biosynthesis in Arabidopsis. Biochemical and Biophysical Research Communications. 2024;695:149423. https://doi.org/10.1016/j.bbrc.2023.149423.

46. Hou L, Chen H, Yan W, Zhang Y, Li M, Wu Q, et al. The VvDREB2A-VvGolS3 regulatory module enhances cold tolerance in Vitis vinifera L. Plant Physiology and Biochemistry. 2025;229:110821. https://doi.org/10.1016/j.plaphy.2025.110821.

47. Nishizawa A, Yabuta Y, Shigeoka S. Galactinol and Raffinose Constitute a Novel Function to Protect Plants from Oxidative Damage. Plant Physiology. 2008;147:1251–63. https://doi.org/10.1104/pp.108.122465.

48. Egert A, Keller F, Peters S. Abiotic stress-induced accumulation of raffinose in Arabidopsis leaves is mediated by a single raffinose synthase (RS5, At5g40390). BMC Plant Biol. 2013;13:218. https://doi.org/10.1186/1471-2229-13-218.

49. Dai H, Zhu Z, Wang Z, Zhang Z, Kong W, Miao M. *Galactinol synthase 1* improves cucumber performance under cold stress by enhancing assimilate translocation. Horticulture Research. 2022;9:uhab063. https://doi.org/10.1093/hr/uhab063.

638 50. Wang S, He P, Wang Z, Zhang H, Meng S, Qi M. Galactinol synthase 4 influences plant
639 height by affecting phenylpropanoid metabolism and the balance of soluble carbohydrates in
640 tomato. Plant Physiology and Biochemistry. 2025;220:109484.
641 https://doi.org/10.1016/j.plaphy.2025.109484.
642 51. Kang M, Wu H, Yang Q, Huang L, Hu Q, Ma T, et al. A chromosome-scale genome
643 assembly of Isatis indigotica, an important medicinal plant used in traditional Chinese
644 medicine: An Isatis genome. Hortic Res. 2020;7:18. https://doi.org/10.1038/s41438-020-
645 0240-5.
646 52. Reiser L, Bakker E, Subramaniam S, Chen X, Sawant S, Khosa K, et al. The Arabidopsis
647 Information Resource in 2024. GENETICS. 2024;227:iyae027.
648 https://doi.org/10.1093/genetics/iyae027.
649 53. Chen R, Liu Q, Cai Q, Qing Y, Wang L, Liu F, et al. Deciphering lipid biosynthesis and
650 regulation in Isatis species through combined metabolomic and transcriptomic analysis. BMC
651 Plant Biol. 2026;26:1139. https://doi.org/10.1186/s12870-026-08944-6.
652 54. Tan Y, Yang J, Jiang Y, Wang J, Liu Y, Zhao Y, et al. Functional Characterization of UDP-
653 Glycosyltransferases Involved in Anti-viral Lignan Glycosides Biosynthesis in Isatis
654 indigotica. Front Plant Sci. 2022;13:921815. https://doi.org/10.3389/fpls.2022.921815.

Figures and figure legends

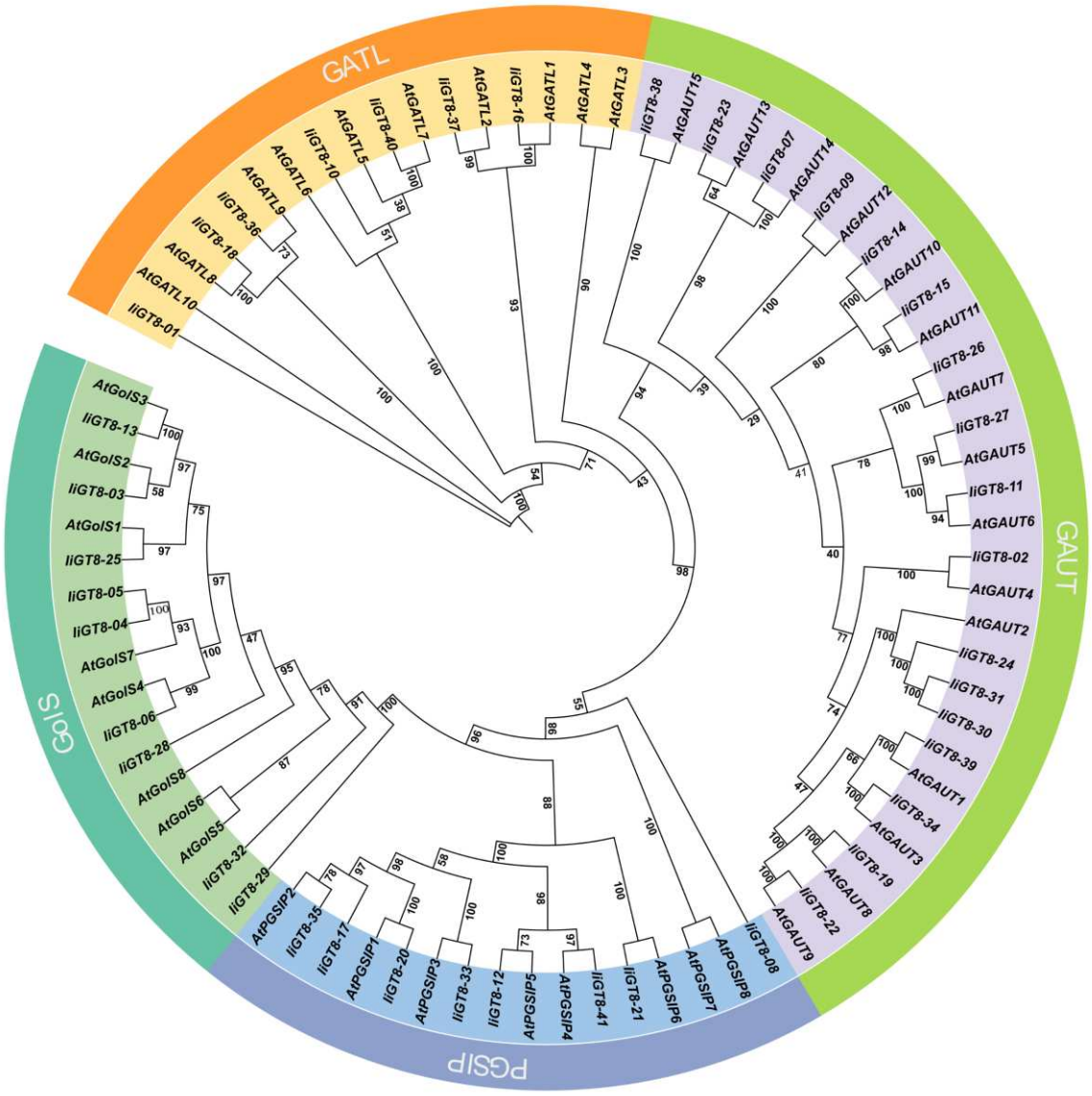


Fig. 1 Phylogenetic relationships of GT8 proteins in *Isatis indigotica* and *Arabidopsis thaliana*. The maximum-likelihood tree includes 41 IiGT8 proteins and functionally characterized AtGT8 proteins. Colored arcs indicate the GAUT, GATL, GolS, and PGSIP subfamilies. Values at nodes are bootstrap percentages based on 1000 replicates.

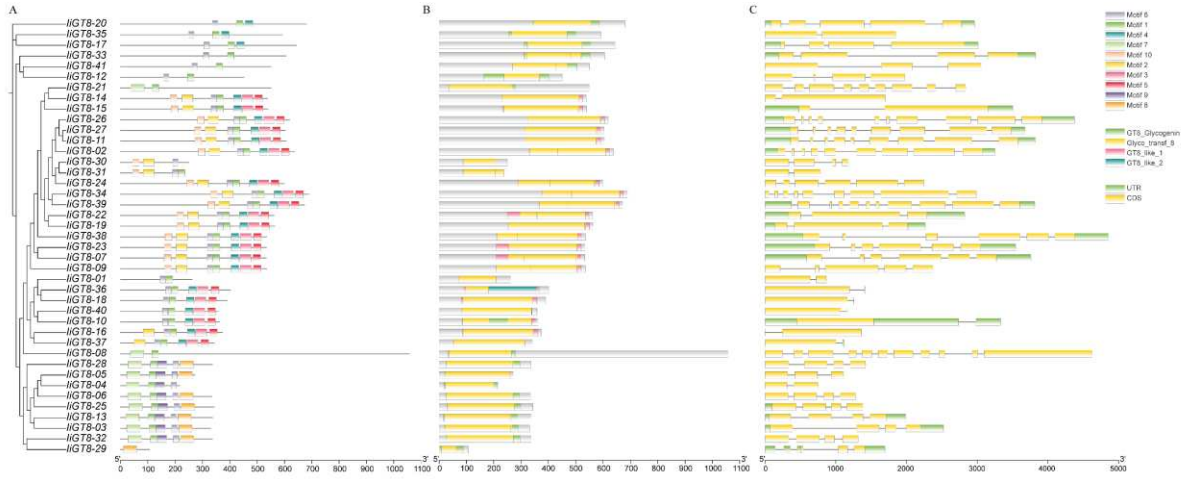


Fig. 2 Conserved motifs, domains, and gene structures of IiGT8 family members. (A) Maximum-likelihood tree and distribution of ten conserved motifs. Colored boxes indicate motifs, and grey lines indicate regions without an assigned motif. (B) Conserved-domain architecture. (C) Exon–intron organization; yellow boxes denote coding sequences, green boxes denote untranslated regions, and black lines denote introns. Scale bars indicate amino-acid length in A and B and nucleotide length in C.

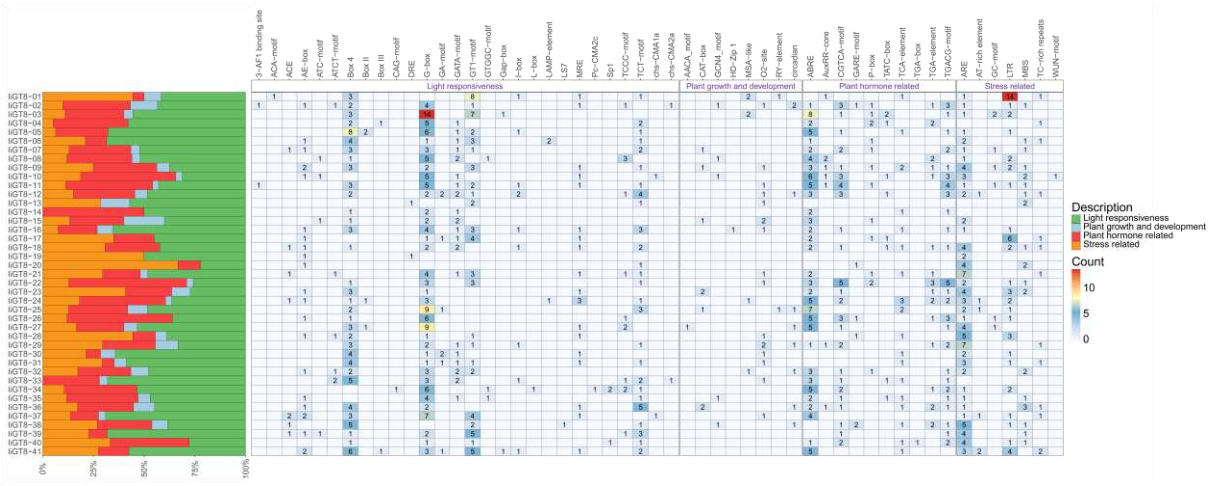
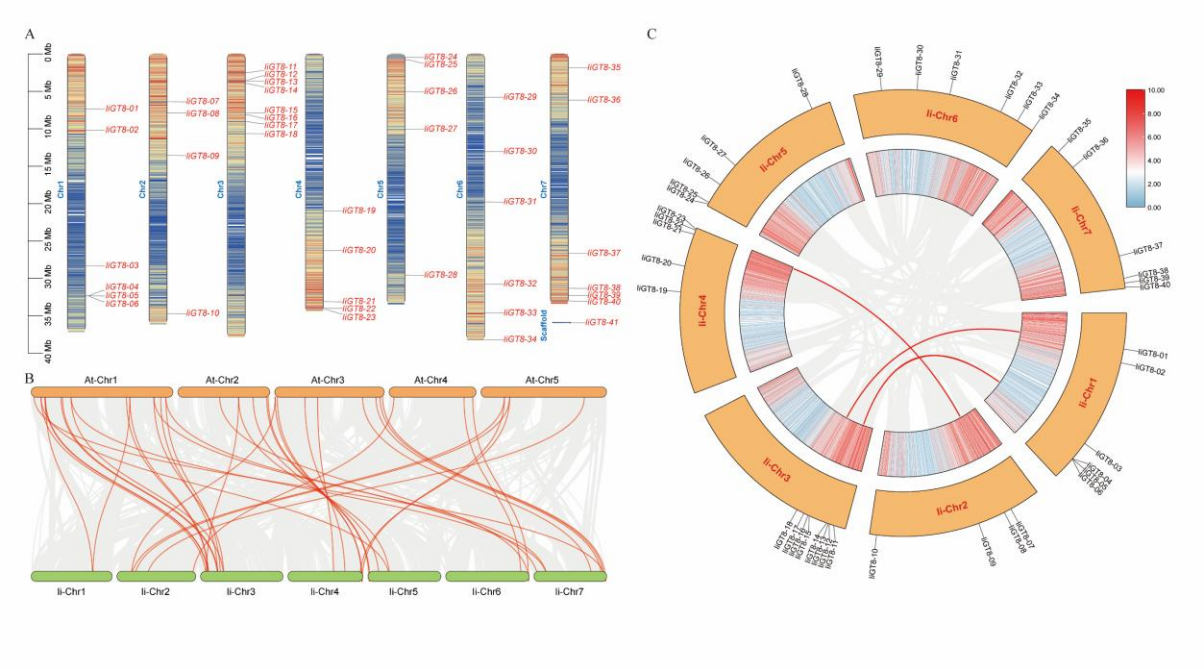


Fig. 3 Distribution of predicted cis-acting elements in IiGT8 promoters. Predicted elements within the 2-kb upstream regions were grouped by functional annotation. Rows represent IiGT8 genes, and columns represent cis-element types. Cell values indicate motif counts, and blank cells indicate zero occurrences. Stacked bars show the proportions of the four displayed functional categories for each gene. ABRE, abscisic acid-responsive element; MBS, MYB-binding site associated with drought inducibility; LTR, low-temperature-responsive element.



675

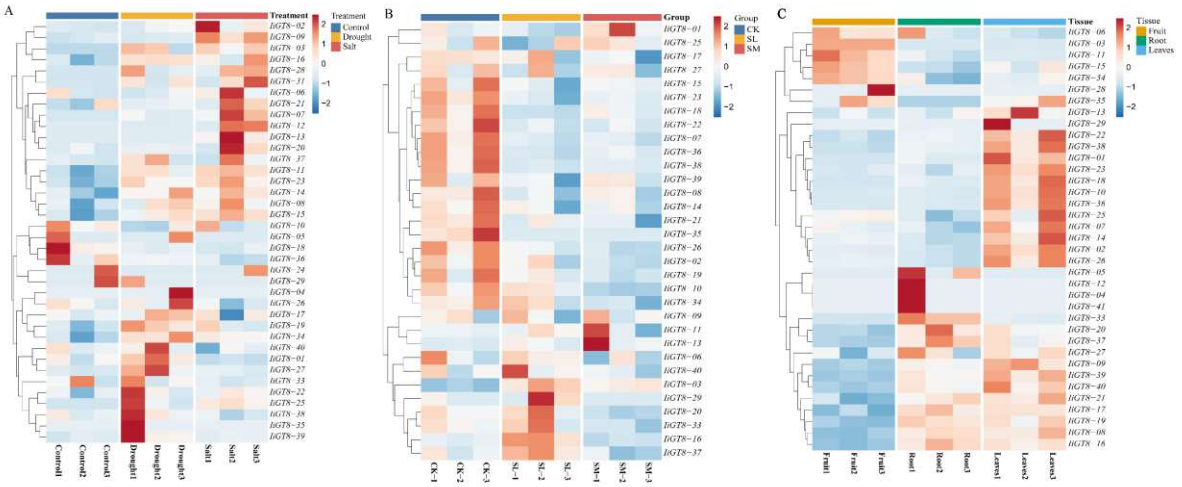
676 **Fig. 4** Chromosomal distribution and synteny of IiGT8 genes. (A) Chromosomal positions of

677 IiGT8 genes on the *Isatis indigotica* chromosomes. (B) Synteny between *Isatis indigotica* and

678 *Arabidopsis thaliana*; gray lines indicate syntenic blocks and red lines highlight GT8 pairs.

679 (C) Intragenomic synteny among *Isatis indigotica* chromosomes; red curves highlight the

680 identified segmentally duplicated IiGT8 pairs.



681

682 **Fig. 5** IiGT8 genes expression profiles under abiotic stress. (A) Expression under control,

683 drought, and salt conditions. (B) Time-course expression under salt treatment (SM, 24 h; SL,

684 48 h). (C) Expression in fruit, roots, and leaves. Colors represent row-wise Z-score-

685 standardized transcript abundance within each dataset. Genes with no detectable expression

686 in the corresponding dataset were omitted.

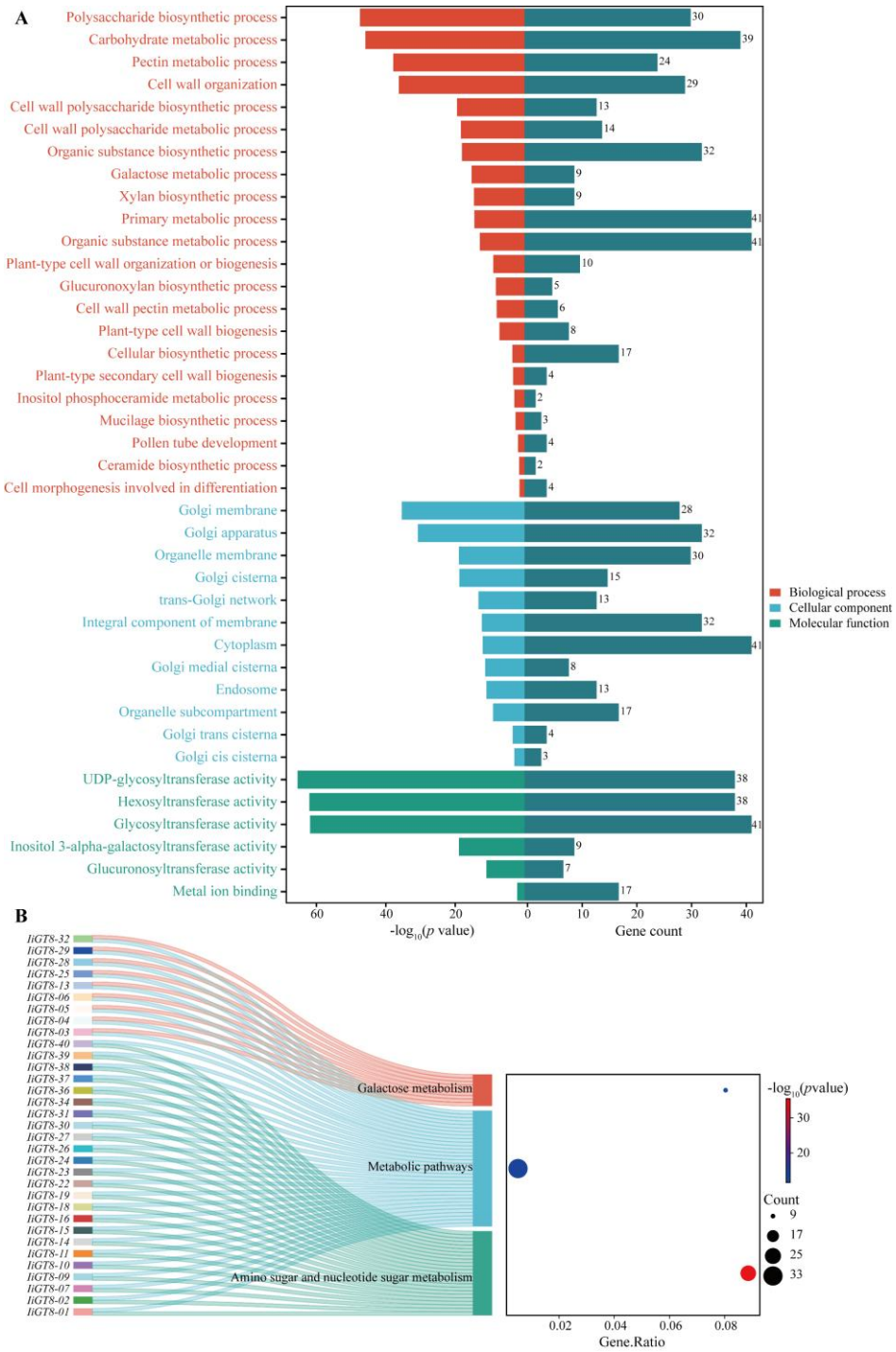
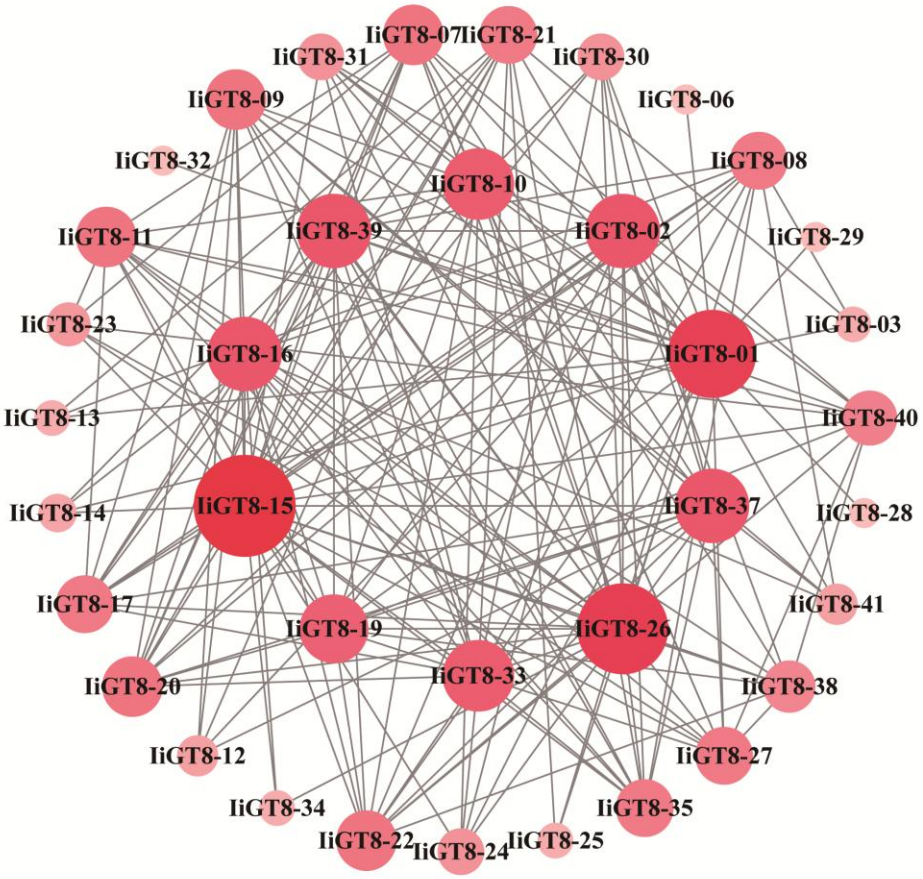


Fig. 6 Functional enrichment of IiGT8 proteins. (A) Gene Ontology enrichment in biological process, molecular function, and cellular component categories. (B) Kyoto Encyclopedia of Genes and Genomes pathway enrichment. The displayed terms summarize the enrichment output; statistical interpretation in the text is restricted to terms with FDR < 0.05. Exact FDR values are provided in Additional file 6: Table S6.



693

694 **Fig. 7** Predicted functional-association network of IiGT8 proteins. The network was

695 generated with STRING and visualized in Cytoscape v3.8.0. Nodes represent IiGT8 proteins,

696 node size reflects connectivity, and edges represent predicted functional associations rather

697 than confirmed physical interactions.

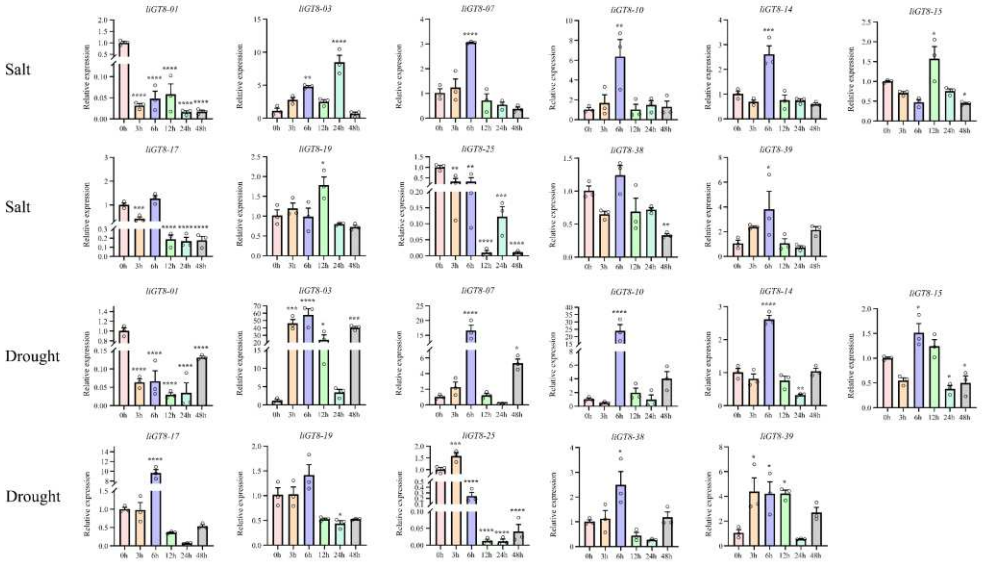


Fig. 8 IiGT8 genes expression under NaCl and PEG treatments. Relative expression was calculated with the $2^{-\Delta\Delta Ct}$ method using the pretreatment 0-h group as the calibrator for each gene and treatment (150 mM NaCl or 20% PEG-6000). Values represent fold expression compared with the expression level at 0 h. Bars show the mean ± SEM, and open circles denote biological replicates (n = 3). For each gene and treatment, statistical differences in the $2^{-\Delta\Delta Ct}$ relative expression values were assessed by one-way ANOVA followed by Dunnett's multiple-comparisons test, comparing each post-treatment time point with the 0-h baseline. Time-matched untreated controls were not included; significance indicates a difference from the 0-h baseline, not a treatment effect relative to an untreated sample collected at the same time. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, and ****$p < 0.0001$.

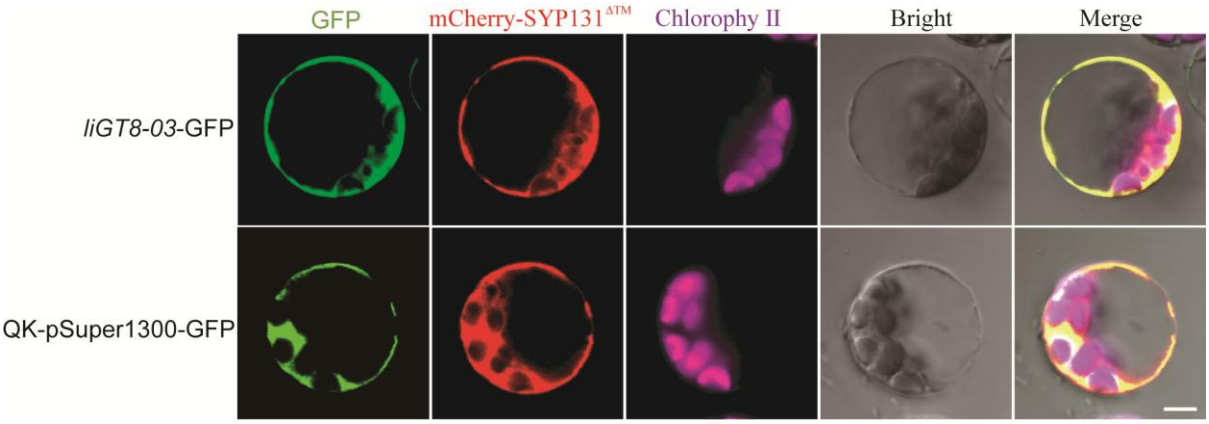


Fig. 9 Subcellular distribution of *IiGT8-03*–GFP in Arabidopsis protoplasts. The upper row shows *IiGT8-03*–GFP, and the lower row shows the QK-pSuper1300-GFP empty-vector control. Channels show GFP, mCherry–SYP131$^{\Delta TM}$, chloroplast autofluorescence, bright field, and merged images. Representative images from three experimental repeats are shown. Scale bar, 5 μm.

Supplementary Files

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

- Additionalfile7.xlsx
- Additionalfile4.xlsx
- Additionalfile6.xlsx
- Additionalfile1.xlsx
- Additionalfile2.xlsx
- Additionalfile3.xlsx
- Additionalfile5.xlsx
- Additionalfile8.xlsx