Received: 9 July 2023 | Revised: 11 September 2023 | Accepted: 25 September 2023

DOI: 10.1111/ppl.14039

ORIGINAL RESEARCH

Physiologia Plantarum

Large-scale analysis of trihelix transcription factors reveals their expansion and evolutionary footprint in plants

Tong Wu[1] | Qihang Yang[1] | Rong Zhou[2] | Tong Yu[1] | Shaoqin Shen[1] | Rui Cao[1] | Xiao Ma[1,3] | Xiaoming Song[1]

[1]School of Life Sciences/Library, North China University of Science and Technology, Tangshan, Hebei, China

[2]Department of Food Science, Aarhus University, Aarhus, Denmark

[3]College of Horticultural Science & Technology, Hebei Normal University Of Science & Technology, Qinhuangdao, Hebei, China

Correspondence
Xiaoming Song, Rui Cao and Xiao Ma, School of Life Sciences/Library, North China University of Science and Technology, Tangshan, Hebei, China.
Email: songxm@ncst.edu.cn; caoruivv@163.com; maxiaoxiaos@sina.com

Funding information
Key Lab. of Nucleic Research, Tangshan, Grant/Award Number: 2022TS003b; National Natural Science Foundation of China, Grant/Award Number: 32172583; Natural Science Foundation for Distinguished Young Scholar of Hebei Province, Grant/Award Number: C2022209010; Natural Science Foundation of Hebei Province, Grant/Award Number: C2021209005

Edited by J-F. Mao

Abstract

The trihelix transcription factor (*TTF*) gene family is an important class of transcription factors that play key roles in regulating developmental processes and responding to various stresses. To date, no comprehensive analysis of the *TTF* gene family in large-scale species has been performed. A cross-genome exploration of its origin, copy number variation, and expression pattern in plants is also unavailable. Here, we identified and characterized the *TTF* gene family in 110 species representing typical plant phylogenetic taxa. Interestingly, we found that the number of *TTF* genes was significantly expanded in *Chara braunii* compared to other species. Based on the available plant genomic datasets, our comparative analysis suggested that the *TTF* gene family likely originated from the GT-1-1 group and then expanded to form other groups through duplication or deletion of some domains. We found evidence that whole-genome duplication/triplication contributed most to the expansion of the *TTF* gene family in dicots, monocots and basal angiosperms. In contrast, dispersed and proximal duplications contributed to the expansion of the *TTF* gene family in algae and bryophyta. The expression patterns of *TTF* genes and their upstream and downstream genes in different treatments showed a functional divergence of *TTF*-related genes. Furthermore, we constructed the interaction network between *TTF* genes and the corresponding upstream and downstream genes, providing a blueprint for their regulatory pathways. This study provided a cross-genome comparative analysis of *TTF* genes in 110 species, which contributed to understanding their copy number expansion and evolutionary footprint in plants.

1 | INTRODUCTION

Transcription factors (TFs) interact with *cis*-elements in the promoter regions of their target genes and belong to a class of DNA-binding proteins (Pei, Li, et al., 2021; Yu, Bai, et al., 2022). Currently, more than 60 TF families have been detected in plants (Zhang et al., 2022). The trihelix TF (*TTF*) gene family is one of the earliest TF families discovered in plants (Xu et al., 2018). The *TTF* gene family has one or two typical trihelix (helix–loop–helix–loop–helix) structures, which determine the specific binding of GT elements (Feng et al., 2019). *TTF* family genes are divided into five subfamilies, GT-1, GT-2, GTγ, SIP1, and SH4, according to the changes in their α-helical domains (Song, Wu, et al., 2016). It is worth noting that GT-2 got its name because it contains two typical trihelix structures (Dehesh et al., 1992; Völz et al., 2018). *TTF* genes have been reported to play multiple regulatory roles in plant growth and development, for example, embryogenesis, flower development, and in response to abiotic and biotic factors (Kaplan-Levy et al., 2012).

A previous study found that the earliest identified GT-1-binding GT element was the promoter of the light-inducible gene *rbcS-3A* (Green et al., 1987). Some other members of the GT-1 subfamily were later discovered in *Arabidopsis thaliana*, *Oryza sativa*, and *Setaria italia* (Kay et al., 1989; Le Gourrierec et al., 1999; Wang, Zhao, et al., 2018).

The Arabidopsis *PETAL LOSS* (*PTL*) gene belongs to the GT-2 family and can regulate the growth of petals and sepals (Lampugnani et al., 2012). In *O. sativa*, *OsGTγ*-1, *OsGTγ*-2, *OsGTγ-3*, and *OsGTγ*-4 may be associated with cold, drought, and salt stress (Fang et al., 2010). The *O. sativa SHAT-TERING1* (*SHA1*) gene, encoding a SH4-type TF, plays an important role in the activation of cell differentiation (Lin et al., 2007). Some *SIP1* genes have been shown to be involved in plant embryonic development, leaf development and cell proliferation in *Solanum lycopersicum* and *A. thaliana* (Barr et al., 2012; Kuromori et al., 2006).

Due to its importance for plant growth and development, the *TTF* gene family has been identified and analyzed in several plants, including *Phyllostachys edulis* (35 *TTF* genes) (Cheng, Xiong, et al., 2019; Wang et al., 2019), *O. sativa* (41 *TTF* genes) (Win et al., 2017; Xiao et al., 2019), *Glycine max* (71 *TTF* genes) (Liu, Zhang, Li, et al., 2020), *Brachypodium distachyon* (27 *TTF* genes) (Wang et al., 2019), *Medicago truncatula* (38 *TTF* genes) (Liu, Zhang, Ma, et al., 2020), *Fagopyrum tataricum* (31 *TTF* genes) (Ma et al., 2019), *Brassica rapa* (52 *TTF* genes) (Luo et al., 2017), *Gossypium arboreum* (52 *TTF* genes) (Mo et al., 2019), *Populus trichocarpa* (56 *TTF* genes) (Wang et al., 2016), *S. lycopersicum* (36 *TTF* genes) (Yu et al., 2018), *S. italia* (27 *TTF* genes) (Wang, Zhao, et al., 2018), and *A. thaliana* (27 *TTF* genes) (Xu et al., 2018). However, there is still a lack of research on the evolutionary function and structure of the *TTF* gene family in large-scale plants. The aims of this study are as follows: (1) to identify and characterize the *TTF* gene family in species representing different plant clades; (2) describe their phylogenetic relationships, origins, and evolutionary footprints; (3) explore the expression patterns of *TTF* genes at different developmental stages, abiotic and biotic stresses based on transcriptome data, and study the function of *TTF* genes; and (4) construct the interaction network between *TTF* genes and their upstream and downstream genes to provide a blueprint for their regulatory pathway. We believe that this comprehensive analysis will contribute to understanding the evolutionary footprint and function of the *TTF* gene family.

2 | RESULTS

2.1 | Identification and characterization of the TTF gene family

We systematically identified *TTF* genes in 110 representative species, including 26 algae, 3 bryophyta, 2 pteridophyta, 3 gymnosperms,

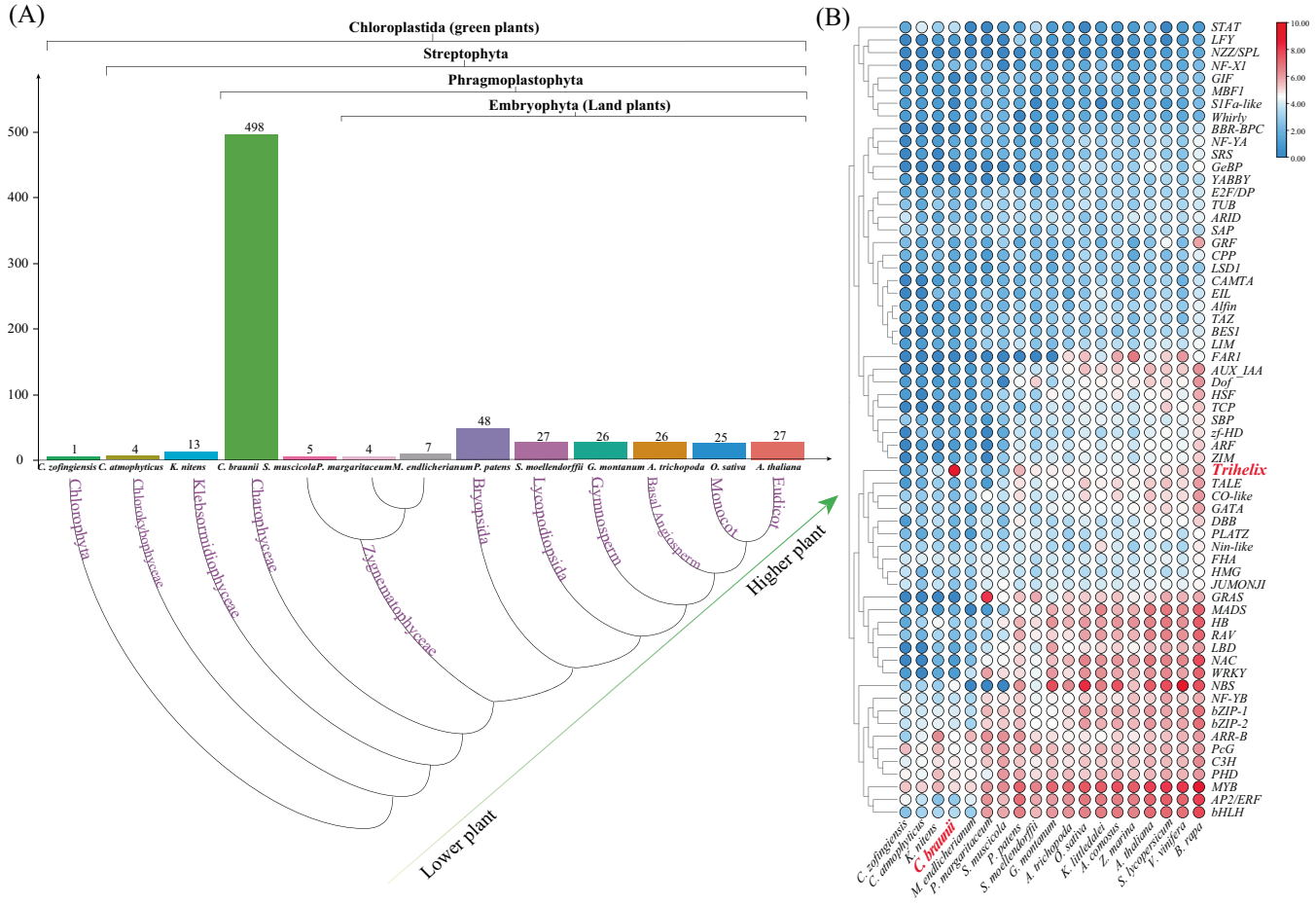

FIGURE 1 Comparative analysis of trihelix transcription factor (*TTF*) genes in representative plants. (A) Phylogenetic tree and number of *TTF* genes for different species in typical taxa. (B) Heatmap of cluster analysis using the number of major transcription factors (TFs) in 18 representative species and *Chara braunii*. The number of TFs was transformed by log 2.

2 basal angiosperms, 8 monocots, and 66 dicots (Table S1). These species represent different branches in the plant phylogenetic tree.

After screening, only 7 of the 26 algae contained *TTF* genes (*Chromochloris zofingiensis*, *Chlorokybus atmophyticus*, *Klebsormidium nitens*, *Chara braunii*, *Mesotaenium endlicherianum*, *Penium margaritaceum*, *Spirogloea muscicola*). According to evolutionary history, as the most primitive algae of these species, *C. zofingiensis*, belonging to the Chlorophyceae, contained only one *TTF* gene (Figure 1A, Table S2). *C. atmophyticus* and *K. nitens* contained 4 and 13 *TTF* genes, respectively (Figure 1A, Table S2). Notably, *C. braunii* belonging to the Charophyceae contained the most *TTF* genes among all species, with the number was 498 (Figure 1A, Table S2). Three more advanced algae were *M. endlicherianum*, *P. margaritaceum*, and *S. muscicola*, which contained seven, four, and five *TTF* genes, respectively (Figure 1A, Table S2). Other land plants encode approximately 30–50 members (Table S2).

Using the same approach, we identified 62 other major TFs in the genomes of *C. braunii* and 18 representative plants (Figure 1B). The 18 species included 4 dicots (*A. thaliana*, *S. lycopersicum*, *Vitis vinifera*, and *B. rapa*); 4 monocots (*O. sativa*, *Ananas comosus*, *Kobresia littledalei*, and *Zostera marina*); 1 basal angiosperm (*Amborella trichopoda*); 1 glymnopsperm (*Gnetum montanum*); 1 pteridophyta (*Selaginella moellendorffii*); 1 bryophyta (*Physcomitrella patens*); and 6 green algae. The number of *TTF* genes in *C. braunii* accounted for 58.4% of its total TF (Figure S1A). We found lower numbers of other TFs per family in *C. braunii* than in land plants, with the *TTF* gene family being an exception (Figure 1B, Figure S1B, Table S3). The results showed that the *TTF* gene family was significantly expanded in *C. braunii* compared to other plants, even all higher plants.

2.2 | Phylogenetic and classification analysis of TTF genes

To classify specific *TTF* genes for each species, we constructed phylogenetic trees between each species examined and *A. thaliana*. The *TTF* genes in most species were classified into five groups (GT-1, GT-2, GTγ, SH4, SIP1) based on bootstrap values and phylogenetic topology (Table S2). However, the GTγ group was absent in algae (Table S2). In the other two lower plants, bryophyta and pteridophyta, their *TTF* genes did not either contain the complete five groups (Table S2). Nevertheless, there were five complete groups of *TTF* genes in all gymnosperms, basal angiosperms, monocots, and dicots (Table S2). This result indicates that the *TTF* genes might have undergone evolutionary differentiation in the structure or function of higher plants. This evolutionary differentiation might be necessary for the development and growth of higher plants. In addition, the Ka/Ks analysis in TTF orthologs and the whole genome orthologs showed that the *TTF* gene family genes have undergone more obvious evolution than the whole genome genes. Therefore, the *TTF* gene family may play an important role in the plant during evolution (Figures S2 and S3).

Phylogenetic analysis showed that the vast majority of *C. braunii* *TTF* genes did not belong to the five previously defined clades (Figure 2A, Figure S4, Table S2). To demonstrate the evolutionary relationship of *TTF* genes, we constructed a comprehensive phylogenetic tree using the selection of 18 representative plants from different taxa (Figure 2B). The results showed that the GT-1 group was divided into 2 distinct subgroups (GT-1-1 and GT-1-2). The *TTF* genes

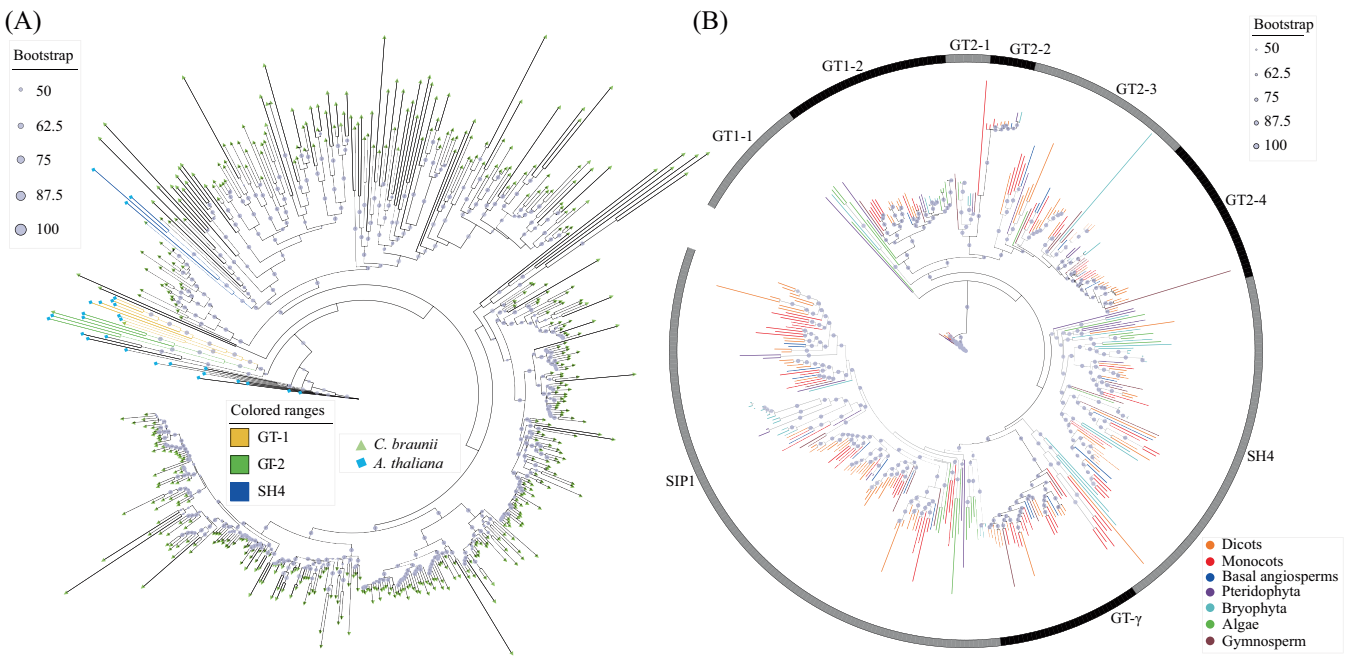

FIGURE 2 Phylogenetic relationship of trihelix transcription factor (*TTF*) gene family. Phylogenetic tree topology was generated via IQTree2. For major nodes, bootstrap values greater than 50% were shown. Different colors indicated groups obtained by bootstrap values and phylogenetic topology. (A) Phylogenetic tree of *Chara braunii*. (B) Phylogenetic tree of 18 representative species.

of *C. zofingiensis* belonged to GT-1-1. The GT-2 group was further divided into four subgroups, corresponding to GT-2-1 to GT-2-4 (Figure 2B).

2.3 | Conserved motif analysis of *TTF* genes to explore origin and evolution pattern

The most conserved 10 motifs (M1–M10) were detected by the MEME program to explore further the origin and evolution of *TTF* genes (Figure 3A, Figure S5). We found specific preservation and expansion of motifs in different groups. As a result, all genes had M1 and M2, while the other eight motifs were only present in some groups (Figure 3A). For example, M3 existed only in GT-2-3 and GT-2-4 groups; M6, M7, and M8 only existed in GT-2-2 group; M9 was only present in the GTγ group, and M10 was only present in the GT-1-2 group (Figure 3A). Of all the groups, only the GT-2 group had two M1. With exception of GT-2-4, the GT-2 group did not have a complete double M2, M4, and M5. In GT-2-3, only M1 corresponded to the trihelix domain. Therefore, we speculated that M1 was an important factor leading to the trihelix.

More importantly, according to the conserved motifs and evolutionary trajectory, we found that the most primitive *TTF* genes of *C. zofingiensis* belonged to GT-1-1 (Figure 3B). Phylogenetic analysis also showed that the *TTF* genes were derived from GT-1-1 (Figures 2A and 3A). This phenomenon indicated that the *TTF* gene family might have originated from the GT-1-1 group and expanded to form other groups by duplication or deletion of some motifs (such as M6–M10). We annotated information on the oldest species present in each group (Figure 3B, Table S2). It was worth noting that GT-1-1, GT-1-2, SH4, SIP1, GT-2-1, and GT-2-3 belonged to primitive algae, while GTγ exclusive, including M9, first appeared in gymnosperms. We speculated that M9 might play an important role in the development and resistance of higher plants.

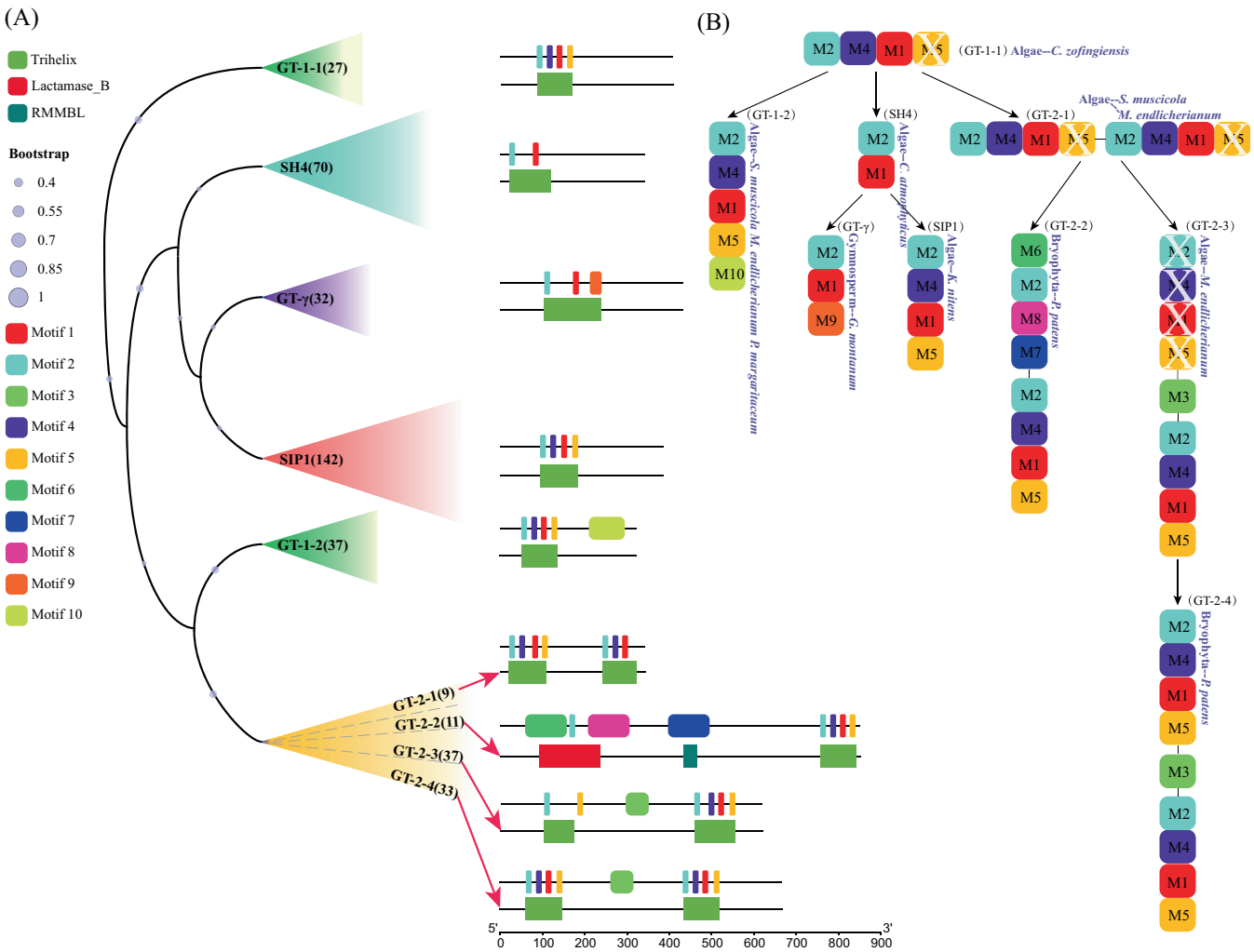

FIGURE 3 Analysis of conserved motifs and evolutionary trajectories of plant trihelix transcription factor (*TTF*) gene family. (A) Motif 1 to motif 10 (M1–M10) of each group of *TTF* gene family. (B) Major evolutionary trajectories of the *TTF* gene family. A white X indicated that the motif was lost in some genes. Annotated species were the oldest species present in each group.

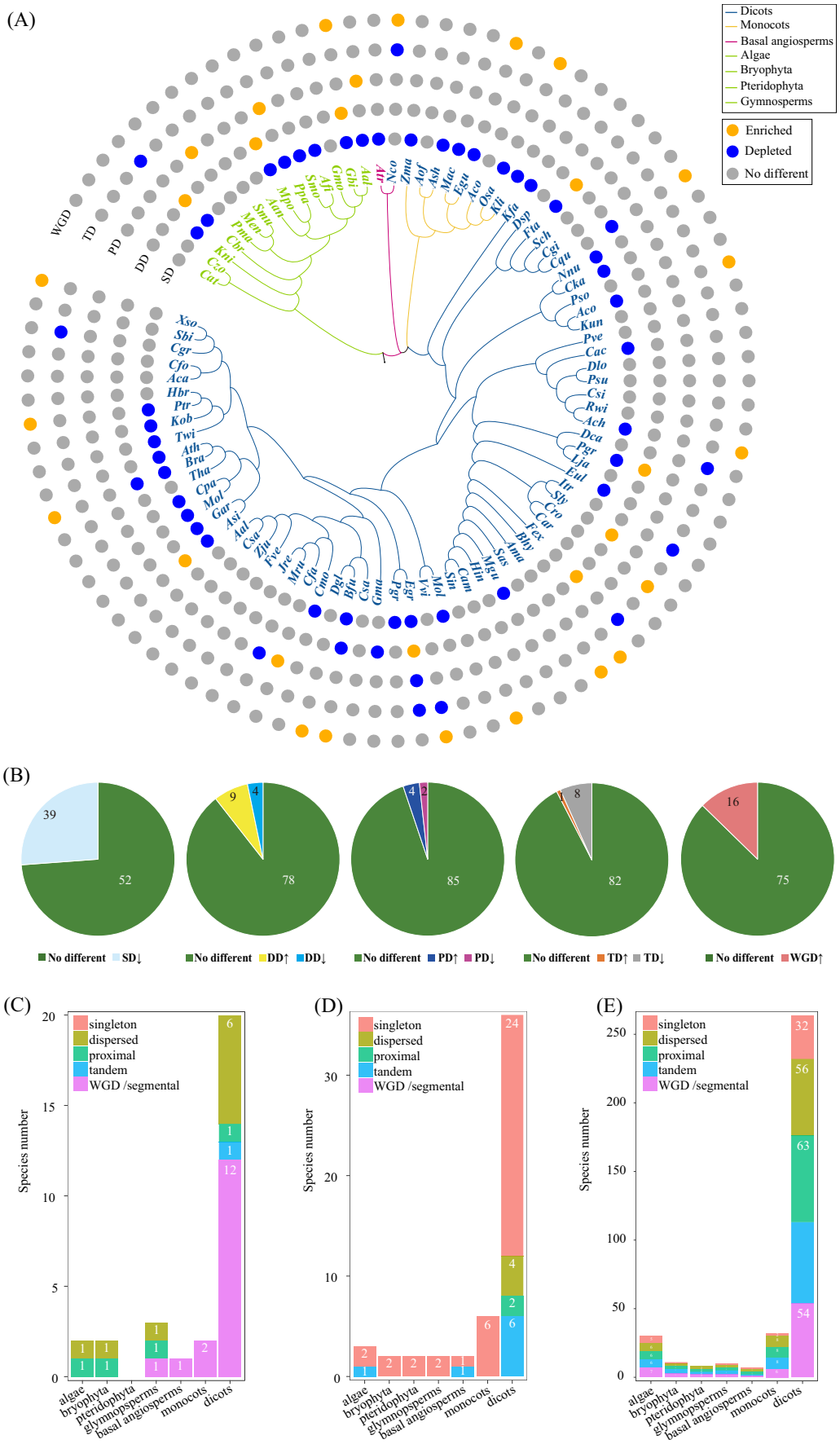

FIGURE 4 Legend on next page.

2.4 | Different expansion mechanisms of *TTF* genes in lower and higher plants

Here, we detected five types of gene duplications, including singleton, dispersed, proximal, tandem, and whole-genome duplication (WGD) or segmental duplication, using the Multiple Collinearity Scan toolkit (MCScanX) program (Figure 4A,B, Tables S4–S8). *TTF* genes were underrepresented (*P*-value <0.05) among the singleton duplication genes in 39 species (Figure 4D, Table S4). Dispersed duplication genes and proximal duplication genes were significantly enriched (*P*-value <0.05) compared with genome-wide levels in algae and bryophyta (Figure 4C, Tables S5 and S6). No duplicated type contributed to the expansion of gene families in pteridophyta (Figure 4C, Tables S4–S8). The no different duplicated type of *TTF* gene numbers in each species taxonomy is also shown (Figure 4E). *TTF* genes were significantly enriched (*P*-value <0.05) in gymnosperms for duplication of dispersed genes, proximal genes, and WGD genes (Figure 4C, Tables S5, S6, and S8). In basal angiosperms and monocots, only WGD genes were significantly enriched for *TTF* genes (*P*-value <0.05) (Figure 4D, Table S8). In 12 of the 66 dicots, *TTF* genes were also enriched (*P*-value <0.05) for WGD genes (Figure 4D, Table S8). WGD accounted for 60% of the dicot gene family expansion. We found evidence that WGD contributed most to the expansion of this gene family in the dicots, monocots and basal angiosperms, while dispersed and proximal duplications contributed to the expansion in algae and bryophyta.

2.5 | Comparative expression pattern analysis of *A. thaliana* and *C. braunii* TTF genes

Their expression at different developmental stages and various stress treatments was compared to examine functional divergence in TTF genes. The normalized expression values underwent log2 conversion. Among the 27 *TTF* genes in *A. thaliana*, 22 were detected in different developmental stages (e.g., embryogenesis, flower development) and responses to abiotic and biotic factors (Figure 5A–C). Most genes in the SIP1 and GT-2 groups showed higher expression at different developmental stages, except *AT2G38250* and *AT5G01380* (Figure 5A, Table S9). Furthermore, the highest expression of *AT1G33240* was observed at flowering, leaf and rosette leaf stages (Figure 5A, Table S9). *AT2G38250* and *AT5G01380* were highly expressed in botrytis cinerea and pseudomonas syringae under biotic stress treatments (Figure 5B, Table S10). A similar result was also found in *S. lycopersicum* (Figure S6). Among abiotic stress treatments, *AT2G38250*, *AT5G01380*, and *AT3G10040* were highly expressed in different treatments, such as cold, osmotic, salt, drought, genotoxic, oxidative, ultraviolet-B, wounding, and heat (Figure 5C, Table S11). The results showed that *AT2G38250*, *AT5G01380*, and *AT3G10040* might be closely related to abiotic and biotic factors. Other *TTF* genes might be involved in growth and development.

The expression data of *C. braunii* involved four tissues, including archegonia, antheridia, zygotes and whole plant (Figure 5D, Table S12). Among the 498 *TTF* genes in *C. braunii*, only 229 genes with the expression values were detected. However, these genes also had distinct expression patterns in the four tissues, with some genes showing tissue-specific expression (Figure 5D). To understand the homology of the *A. thaliana* and *C. braunii TTF* genes, Blastp was used to find the best match (*E*-value <1×10^{-5}, identity >40%). After filtering, a total of 22 *C. braunii* genes best-matched *A. thaliana* (Figure 5D, Table S13). As a result, only genes in the SIP1 group did not have any best matches (Figure 5D, Table S13). Notably, the best match for gene *g29889* (the gene ID in the *C. braunii* genome) was *AT1G33240* (GT-2), and the gene pair was more than 50% identical (Table S13). Furthermore, *g29889* was highly expressed in all four tissues (Figure 5d). Meanwhile, *AT1G33240* (GT-2) had the highest expression across the entire developmental map (Figure 5D). This result might indicate that the two genes might have similar functions. Based on the homologous gene relationship with *Arabidopsis*, it will help us to further explore the function of corresponding genes in *C. braunii*.

2.6 | Construction of regulatory network to explore *TTF* gene functions

Comparative analysis of expression patterns and construction of regulatory networks will provide very favorable support for studying trihelix gene functions in *A. thaliana* and other related species. First, we obtained the upstream and downstream genes of the *A. thaliana TTF* gene family using the iGRN database. Then, an interaction network was constructed between each *TTF* gene and the corresponding upstream and downstream genes to uncover their regulatory pathways (Figures S7–S10, Tables S14–S18). For example, we detected 580 upstream and 1245 downstream regulatory interactions in the GT-2 group (Figure 6A,B, Table S14). The green dots represented all genes, and the *TTF* genes have been marked. The red line represents the upstream regulatory genes of *TTF*, and the green link represents the downstream regulatory genes of *TTF* genes. Furthermore, we conducted enrichment analysis (*P*-value <0.05) for the families of upstream and downstream genes involved in the network. Among the downstream genes, no common enriched family existed for each

FIGURE 4 Plant phylogeny and enrichment analysis of each duplicated type of trihelix transcription factor (*TTF*) genes. (A) Each duplicated type is shown with a solid circle, and the significance levels relative to the genome-wide mean were shown in different colors, enriched (orange, $P < 0.05$), depleted (blue, $P < 0.05$), or not significantly difference (gray). (B) Proportion of each significantly enriched or depleted duplicated type in total copy type. (C) Species number of enriched duplicated type of *TTF* genes in each species taxonomy. (D) Species number of depleted duplicated type of *TTF* genes in each species taxonomy. (E) Species number of no different duplicated type of *TTF* genes in each species taxonomy.

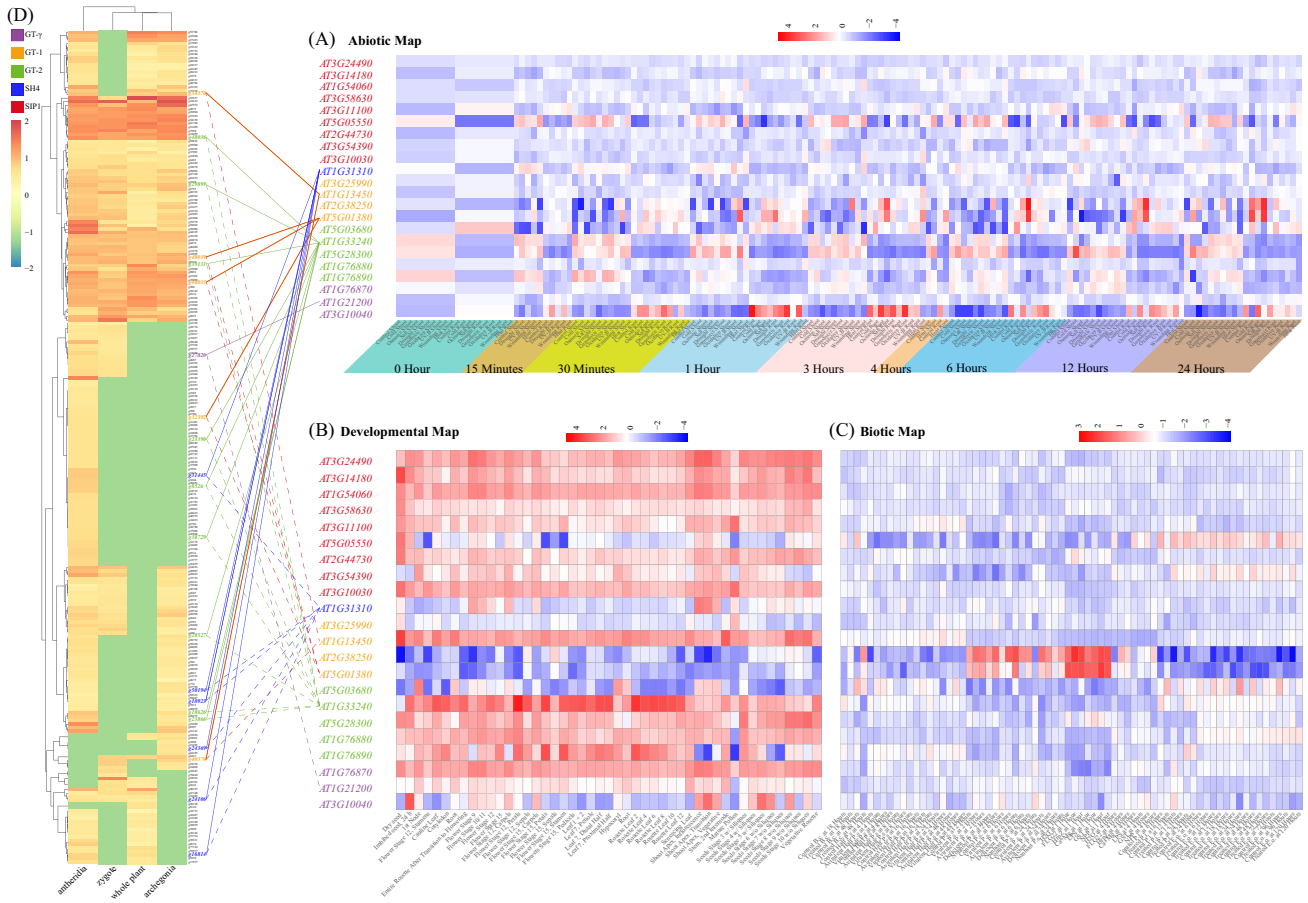

FIGURE 5 Comparative expression pattern analysis of trihelix transcription factor (*TTF*) genes in *Arabidopsis thaliana* and *Chara braunii*. (A) Expression patterns of *TTF* genes in *A. thaliana* under abiotic stress. (B) Expression patterns of the *A. thaliana TTF* genes in the developmental map. (C) Expression patterns of *TTF* genes in *A. thaliana* under biotic stress. (D) Expression patterns of *C. braunii TTF* genes in four tissues.

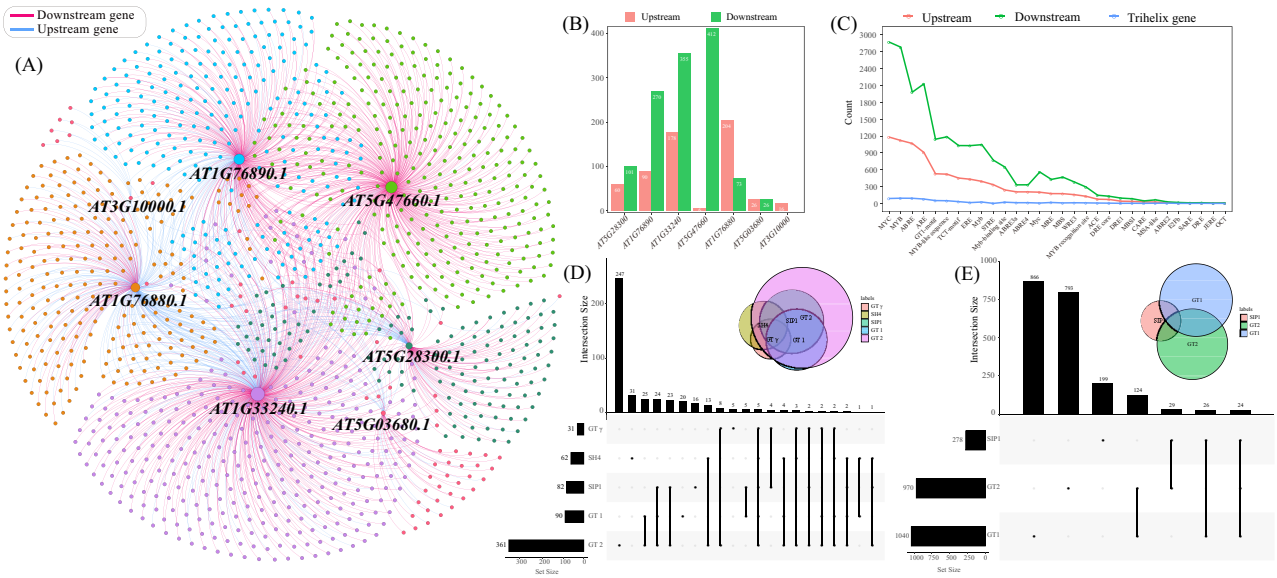

FIGURE 6 Interaction network analysis of trihelix transcription factor (*TTF*) genes and their upstream and downstream genes in GT-2. (A) The interaction network between *TTF* genes and their related genes was constructed according to iGRN. (B) Number of genes with upstream and downstream regulatory interactions for each *TTF* gene. (C) *Cis*-acting element analysis of upstream, downstream genes, and *TTF* genes. (D) Venn diagram representing the number of common or specific upstream genes in each group. (E) Venn diagram representing the number of common or specific downstream genes in each group.

group. Among upstream genes, SRF-TF was a common enriched family in all groups (Figure S11, Tables S19–S23). AP2 was enriched in GT-2, GT-1, SIP1, and SH4 (Figure S11, Tables S19–S22). Each group also had unique enriched families (Figure S11, Tables S19–S23). These results indicated that they might play a more central role in regulatory networks. In the analysis of *cis*-acting elements of all related genes, we found that they had more MYC, MYB, ABRE, ARE, and ERE elements in the upstream and downstream genes (Figure 6C, S4C, S5C, S6C, S7C, Tables S24–S28). Most of these *cis*-acting elements were related to resistance, indicating that *TTF* genes were closely related to resistance.

We drew a Venn diagram of the upstream and downstream genes of these five groups of the *TTF* gene family (Figure 6D,E). Among the five groups of upstream genes, there were a total of 5 genes (*AT5G13790*, *AT1G69120*, *AT3G54340*, *AT1G24260*, *AT5G20240*) (Table S29). We found that five common upstream genes were also associated with plant development and responses to abiotic and biotic factors according to the expression pattern analysis (Figure S12). Expression pattern analysis showed that *AT5G20240* was highly expressed across the entire abiotic stress profile (Figure S12A, Table S30). *AT1G69120*, *AT3G54340*, *AT1G24260*, and *AT5G20240* had higher expression during flower development (Figure S12B, Table S31). Especially, *AT3G54340* and *AT1G24260* were also expressed at the seed stage (Figure S12B, Table S31). *AT1G69120*, *AT3G54340*, *AT1G24260*, and *AT5G20240* had higher expression across the entire biotic stress profile (Figure S12C, Table S32).

Among the downstream genes, there were a total of 24 genes in common (Figure S13, Table S33). Among these 24 genes, 22 genes were detected to be expressed. Expression pattern analysis showed that genes in group2 had the highest expression across the entire abiotic stress profile (Figure S13A, Table S34). Genes of group1 and group2 were highly expressed on the developmental map (Figure S13B, Table S35). In addition, the genes of group2 were highly expressed on the biotic stress map (Figure S13C, Table S36). All these phenomena indicated that the most common genes might be related to their common biological functions.

3 | DISCUSSION

The primitive ancestor of green plants came into contact with land about 500 million years ago and is thought to have evolved from a class of algae in streptophyta (Virtanen et al., 2020). The diversity of streptophyta is very large, of which *C. braunii* is considered to be the most closely related to terrestrial plants (Martin & Allen, 2018; Wang, Li, et al., 2020). *C. braunii* is the only algae with a tissue-like structure that can differentiate into root-like protrusions anchored on solid substrates (Nishiyama et al., 2018). Furthermore, *C. braunii* has a unique xylan synthase for cell wall biosynthesis, a phragmoplast (cell separation) mechanism similar to land plants and many phytohormones (Beilby, 2019; Nishiyama et al., 2018). Similar to land plants, *C. braunii* plastids are controlled by retrograde signals, with more refined transcriptional regulation than other algae (Bonnot et al., 2019). *TTF* genes have been reported to play multiple regulatory roles in plant growth and development, as well as response to abiotic and biotic factors (Song, Wu, et al., 2016; Wang et al., 2017).

In our study, most algae contained *TTF* genes belonging to streptophyta. Among them, *C. braunii* contained the most *TTF* genes (498) in all lower and higher plants. Moreover, over 50% of TFs in *C. braunii* belonged to the *TTF* gene family. Therefore, we speculated that the expansion of the *TTF* genes in *C. braunii* might be related to the independent evolution of its morphological complexity. In addition, depending on the function of *TTF* genes, they might provide a solid foundation for species responses to abiotic and biotic factors during the transformation of aquatic plants to terrestrial plants. Interestingly, our analysis showed that more than half of the *TTF* genes in *C. braunii* had no detectable expression using available RNA-seq datasets. This phenomenon indicated that although the *TTF* gene family had a mass outbreak in number, many genes still did not play their corresponding functions. This phenomenon was also consistent with previous reports that duplicated genes were neo-functionalization, sub-functionalization, and lost to reduce gene redundancy (Birchler & Yang, 2022).

Duplicated type identification analysis of the *TTF* family indicated that the expansion and evolution mechanisms were different in lower and higher plants (Song et al., 2020). Previous studies have shown that in the absence of WGD, gene family expansions resulted from gene duplication and differential loss for most lower plants (Adams & Wendel, 2005; Rieseberg et al., 2003). In lower plants, we found that dispersed duplication and proximal duplication played a greater role in the expansion of the *TTF* genes. The *TTF* genes of *C. braunii* were significantly enriched for dispersed duplication genes at the genome-wide level, which directly contributed to the expansion of its *TTF* genes. But for *S. muscicola*, all 5 *TTF* genes were derived from WGD because it underwent a whole-genome triplication, according to a previous report (Cheng, Xian, et al., 2019).

WGD/T played a greater role in the expansion of *TTF* genes in higher plants. These findings suggested that expansion of the *TTF* genes in a wide range of higher plants might be associated with polyploidization. Many WGD/T events were distributed in angiosperms and might lead to gene expression and epigenetic remodeling changes, which might provide variability that allowed rapid adaptation to new environments (Hegarty & Hiscock, 2007; Taylor & Raes, 2004).

Genes containing the conserved domain of SRF-TF were called MADS-box genes, which were involved in plant floral organ development and regulation of flowering time (Wang, Chen, et al., 2018). Among upstream genes, SRF-TF was a common enriched family in each group. In our analysis of expression patterns in *A. thaliana*, some genes were highly expressed in the developmental map. This result was consistent with previous studies (Kaplan-Levy et al., 2014; Xu et al., 2018). The expression of these genes in different developmental stages and tissues could help us understand their specific functions. Genes containing conserved domains of AP2 have been implicated in the induction of various physiological and biochemical signals, such as disease and stress resistance (Feng et al., 2020; Song et al., 2013; Song, Wang, et al., 2016). Among upstream genes, AP2 was enriched in GT-2, GT-1, SIP1, and SH4. In conclusion, we preliminarily explored the function of *TTF* genes based on gene expression data of Arabidopsis and *C. braunii*.

4 | MATERIALS AND METHODS

4.1 | Retrieval of genome sequences

We collected genome sequences from 110 plants representing different branches in the plant phylogenetic tree according to plaBiPD (https://www.plabipd.de/plant_genomes_pn.ep), TVIR database (http://tvir.bio2db.com) (Yu, Ma, et al., 2022), and TBGR database (http://www.tbgr.org.cn) (Liu et al., 2022). These species included 26 algae, 3 bryophyta, 2 pteridophyta, 3 gymnosperms, 2 basal angiosperms, 8 monocots, and 66 dicots. Genome-related datasets for these species were downloaded from several databases, such as NCBI (https://www.ncbi.nlm.nih.gov), Phytozome (https://phytozome-next.jgi.doe.gov), Ensembl Plants (http://plants.ensembl.org/index.html), and other related databases (Table S1).

4.2 | Identification and characterization of the *TTF* gene family

To identify putative *TTF* family members, the hidden Markov model profile of *TTF* (PF13837) was obtained from the Pfam (version 35.0) database (http://pfam.xfam.org/) (El-Gebali et al., 2019). Then, it was used to identify the putative *TTF* gene family with a threshold of E-value <1e-5 (El-Gebali et al., 2019; Pei, Yu, et al., 2021). The retrieved *TTF* candidates were further validated by SMART and conserved domain database with a threshold of E-value <1e-5 (Letunic et al., 2012; Marchler-Bauer et al., 2015).

4.3 | Sequence alignment and phylogenetic analysis

Multiple sequence alignments were performed using MUSCLE (version 3.8.31) with default parameters (Edgar, 2004). Based on the sequence alignment, the neighbor-joining method was used to construct the *TTF* gene phylogenetic tree by IQTree2 (version 1.6.2), and the Bootstrap value was set to 1000 (Minh et al., 2020). Single-copy ortholog sequences were identified using OrthoFinder (v2.2.7, https://github.com/davidemms/OrthoFinder/releases) to construct species trees (Emms & Kelly, 2019). MEME (version 5.3.3) was used to search for conservative motifs, and the number of motifs was set to 10 (Bailey et al., 2009).

4.4 | Identification of gene collinearity and specific duplication events

MCScanX was used for gene collinearity with default parameters, according to a previous report (Wang et al., 2012). A program (duplicate_gene_classifier) in MCScanX was used to infer different types of duplicated genes (Song, Sun et al., 2021; Song, Wei et al., 2021). We extracted *TTF* genes located in collinear blocks by Perl script. Then, chi-square test was used to determine whether the *TTF* gene family was significantly enriched (*P*-value <0.05) in a specific duplication event.

4.5 | TTF gene expression analysis of *C. braunii* and *A. thaliana*

We performed *TTF* gene expression analysis in *C. braunii* using RNA-seq data reported previously (Nishiyama et al., 2018). These data were obtained from four tissues, including archegonia, antheridia, zygotes, and whole plants of *C. braunii*. Fragment per kilobase of transcript per million fragments mapped values were used to normalize gene expression. The expression analysis data of *A. thaliana TTF* genes under abiotic stress treatment, biotic stress treatment and developmental map were extracted from eFP browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi) (Adams & Wendel, 2005). The expression data include biotic stress and abiotic stress. Biological stress includes insect pests and bacterial infection. Abiotic stresses include drought, flood, salinity, mineral deficiency, and adverse pH. Data were normalized by the Genetic Counseling Outcome Scale method (Grant et al., 2019). The heatmap package (https://cran.r-project.org/web/packages/pheatmap/index.html) of R was used to draw expression heatmap (Wang, Hu, et al., 2020).

4.6 | Construction of regulatory networks in *A. thaliana*

The upstream and downstream genes of the *A. thaliana TTF* genes were derived from iGRN (http://bioinformatics.psb.ugent.be/webtools/iGRN/) (De Clercq et al., 2021). The interaction network of these genes was constructed using Gephi software (https://gephi.org) (Amith et al., 2019). Venn diagram (http://bioinformatics.psb.ugent.be/webtools/Venn/) was used to show their overlapping relationship according to previous report (Song, Li, et al., 2021).

4.7 | The *cis*-acting elements analysis of *TTF*-related genes in *A. thaliana*

To illustrate the functions of upstream and downstream genes, TBtools software was used to take the translation initiation codon of the upstream 2Kb (Chen et al., 2020). In addition, PlantCare (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to identify *cis*-acting elements 2Kb upstream of the translation initiation codon (Lescot et al., 2002). Pfam (version 34.0) was used to identify families of these genes (Finn et al., 2014). Furthermore, the Python script was used to perform enrichment analysis (Virtanen et al., 2020). The *P*-values obtained by the significance analysis were further corrected using the Bonferroni method of the R program (Virtanen et al., 2020). Corrected *P*-values (*q*-values) <0.05 and fold-changes >2 were used to define significant enrichment terms according to previous report (Song, Hu, et al., 2021).

AUTHOR CONTRIBUTIONS

Xiaoming Song conceived the project and was responsible for the project initiation. Data generation and bioinformatics analysis were led by Tong Wu, Xiao Ma, Rui Cao, Xiaoming Song, Qihang Yang, Tong Yu, and Shaoqin Shen The manuscript was organized, written and revised by Xiaoming Song, Xiao Ma, Rui Cao, Tong Wu, and Rong Zhou. All authors read and revised the manuscript.

ACKNOWLEDGEMENTS

This work was supported by the Natural Science Foundation of Hebei (C2021209005), the National Natural Science Foundation of China (32172583), the Natural Science Foundation for Distinguished Young Scholar of Hebei Province (C2022209010), and Key Lab. of Nucleic Research, Tangshan (2022TS003b).

CONFLICT OF INTEREST STATEMENT

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

All materials and related data in this study are provided in the supplementary files. The Illumina RNA-seq dataset of *C. braunii* analyzed during the current study was downloaded from the NCBI with the SRA accession number PRJNA445548. Data for the *A. thaliana* expression dataset was obtained from eFP browser (http://bar.utoronto.ca/efp_arabidopsis/cgi-bin/efpWeb.cgi).

ORCID

Xiaoming Song https://orcid.org/0000-0003-0084-3668

REFERENCES

Adams, K. & Wendel, J. (2005) Novel patterns of gene expression in polyploid plants. *Trends in Genetics*, 21(10), 539–543.

Amith, M.T., Fujimoto, K. & Tao, C., (2019). NET-EXPO: a gephi plugin towards social network analysis of network exposure for unipartite and bipartite graphs. In: *HCI International 2019-Posters: 21st International Conference, HCII 2019, Orlando, FL, USA, July 26-31, 2019, Proceedings, Part III 21* (pp. 3–12). Springer International Publishing.

Bailey, T., Boden, M., Buske, F., Frith, M., Grant, C.E., Clementi, L. et al. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server issue), W202–W208.

Barr, M., Willmann, M.R. & Jenik, P.D. (2012) Is there a role for trihelix transcription factors in embryo maturation? *Plant Signaling & Behavior*, 7(2), 205–209.

Beilby, M. (2019) *Chara braunii* genome: a new resource for plant electrophysiology. *Biophysical Reviews*, 11(2), 235–239.

Birchler, J. & Yang, H. (2022) The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell*, 34(7), 2466–2474.

Bonnot, C., Hetherington, A.J., Champion, C., Breuninger, H., Kelly, S. & Dolan, L. (2019) Neofunctionalisation of basic helix-loop-helix proteins occurred when embryophytes colonised the land. *The New Phytologist*, 223(2), 993–1008.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. et al. (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant*, 13(8), 1194–1202.

Cheng, S., Xian, W., Fu, Y., Marin, B., Keller, J., Wu, T. et al. (2019) Genomes of subaerial Zygnematophyceae provide insights into land plant evolution. *Cell*, 179(5), 1057–1067.e1014.

Cheng, X., Xiong, R., Yan, H., Gao, Y., Liu, H., Wu, M. et al. (2019) The trihelix family of transcription factors: functional and evolutionary analysis in Moso bamboo (*Phyllostachys edulis*). *BMC Plant Biology*, 19(1), 154.

De Clercq, I., Van de Velde, J., Luo, X., Liu, L., Storme, V., Van Bel, M. et al. (2021) Integrative inference of transcriptional networks in Arabidopsis yields novel ROS signalling regulators. *Nature Plants*, 7(4), 500–513.

Dehesh, K., Hung, H., Tepperman, J.M. & Quail, P.H. (1992) GT-2: a transcription factor with twin autonomous DNA-binding domains of closely related but different target sequence specificity. *The EMBO Journal*, 11(11), 4131–4144.

Edgar, R. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*, 5, 113.

El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C. et al. (2019) The Pfam protein families database in 2019. *Nucleic Acids Research*, 47(D1), D427–D432.

Emms, D. & Kelly, S. (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology*, 20(1), 238.

Fang, Y., Xie, K., Hou, X., Hu, H. & Xiong, L. (2010) Systematic analysis of GT factor family of rice reveals a novel subfamily involved in stress responses. *Molecular Genetics and Genomics*, 283(2), 157–169.

Feng, C., Song, X. & Tang, H. (2019) Molecular cloning and expression analysis of GT-2-like genes in strawberry. *3 Biotech*, 9(3), 105.

Feng, K., Hou, X.L., Xing, G.M., Liu, J.X., Duan, A.Q., Xu, Z.S. et al. (2020) Advances in AP2/ERF super-family transcription factors in plant. *Critical Reviews in Biotechnology*, 40(6), 750–776.

Finn, R., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R. et al. (2014) Pfam: the protein families database. *Nucleic Acids Research*, 42(Database issue), D222–D230.

Grant, P., Pampaka, M., Payne, K., Clarke, A. & McAllister, M. (2019) Developing a short-form of the genetic counselling outcome scale: the genomics outcome scale. *European Journal of Medical Genetics*, 62(5), 324–334.

Green, P., Kay, S.A. & Chua, N.H. (1987) Sequence-specific interactions of a pea nuclear factor with light-responsive elements upstream of the rbcS-3A gene. *The EMBO Journal*, 6(9), 2543–2549.

Hegarty, M. & Hiscock, S. (2007) Polyploidy: doubling up for evolutionary success. *Current Biology*, 17(21), R927–R929.

Kaplan-Levy, R.N., Brewer, P.B., Quon, T. & Smyth, D.R. (2012) The trihelix family of transcription factors—light, stress and development. *Trends in Plant Science*, 17(3), 163–171.

Kaplan-Levy, R.N., Quon, T., O'Brien, M., Sappl, P.G. & Smyth, D.R. (2014) Functional domains of the PETAL LOSS protein, a trihelix transcription factor that represses regional growth in *Arabidopsis thaliana*. *The Plant Journal*, 79(3), 477–491.

Kay, S., Keith, B., Shinozaki, K., Chye, M.L. & Chua, N.H. (1989) The rice phytochrome gene: structure, autoregulated expression, and binding of GT-1 to a conserved site in the 5′ upstream region. *Plant Cell*, 1(3), 351–360.

Kuromori, T., Wada, T., Kamiya, A., Yuguchi, M., Yokouchi, T., Imura, Y. et al. (2006) A trial of phenome analysis using 4000 Ds-insertional mutants in gene-coding regions of Arabidopsis. *The Plant Journal*, 47(4), 640–651.

Lampugnani, E., Kilinc, A. & Smyth, D.R. (2012) PETAL LOSS is a boundary gene that inhibits growth between developing sepals in *Arabidopsis thaliana*. *The Plant Journal*, 71(5), 724–735.

Le Gourrierec, J., Li, Y.F. & Zhou, D.X. (1999) Transcriptional activation by Arabidopsis GT-1 may be through interaction with TFIIA-TBP-TATA complex. *The Plant Journal*, 18(6), 663–668.

Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y. et al. (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 30(1), 325–327.

Letunic, I., Doerks, T. & Bork, P. (2012) SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research*, 40-(Database issue), D302–D305.

Lin, Z., Griffith, M.E., Li, X., Zhu, Z., Tan, L., Fu, Y. et al. (2007) Origin of seed shattering in rice (*Oryza sativa* L.). *Planta*, 226(1), 11–20.

Liu, W., Zhang, Y., Li, W., Lin, Y., Wang, C., Xu, R. et al. (2020) Genome-wide characterization and expression analysis of soybean trihelix gene family. *PeerJ*, 8, e8753.

Liu, X., Zhang, H., Ma, L., Wang, Z. & Wang, K. (2020) Genome-wide identification and expression profiling analysis of the trihelix gene family under abiotic stresses in *Medicago truncatula*. *Genes (Basel)*, 11(11), 1389.

Liu, Z., Li, N., Yu, T., Wang, Z., Wang, J., Ren, J. et al. (2022) The Brassicaceae genome resource (TBGR): a comprehensive genome platform for Brassicaceae plants. *Plant Physiology*, 190, 226–237.

Luo, J., Tang, S., Mei, F., Peng, X., Li, J., Li, X. et al. (2017) BnSIP1-1, a trihelix family gene, mediates abiotic stress tolerance and ABA signaling in *Brassica napus*. *Frontiers in Plant Science*, 8, 44.

Ma, Z., Liu, M., Sun, W., Huang, L., Wu, Q., Bu, T. et al. (2019) Genome-wide identification and expression analysis of the trihelix transcription factor family in tartary buckwheat (*Fagopyrum tataricum*). *BMC Plant Biology*, 19(1), 344.

Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y. et al. (2015) CDD: NCBI's conserved domain database. *Nucleic Acids Research*, 43(Database issue), D222–D226.

Martin, W. & Allen, J.F. (2018) An algal greening of land. *Cell*, 174(2), 256–258.

Minh, B., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A. et al. (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534.

Mo, H., Wang, L., Ma, S., Yu, D., Lu, L., Yang, Z. et al. (2019) Transcriptome profiling of *Gossypium arboreum* during fiber initiation and the genome-wide identification of trihelix transcription factors. *Gene*, 709, 36–47.

Nishiyama, T., Sakayama, H., de Vries, J., Buschmann, H., Saint-Marcoux, D., Ullrich, K.K. et al. (2018) The Chara genome: secondary complexity and implications for plant terrestrialization. *Cell*, 174(2), 448–464.e424.

Pei, Q., Li, N., Bai, Y., Wu, T., Yang, Q., Yu, T. et al. (2021) Comparative analysis of the TCP gene family in celery, coriander and carrot (family Apiaceae). *Vegetable Research*, 1, 5–12.

Pei, Q., Yu, T., Wu, T., Yang, Q., Gong, K., Zhou, R. et al. (2021) Comprehensive identification and analyses of the Hsf gene family in the whole-genome of three Apiaceae species. *Horticultural Plant Journal*, 7(5), 457–468.

Rieseberg, L.H., Raymond, O., Rosenthal, D.M., Lai, Z., Livingstone, K., Nakazato, T. et al. (2003) Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, 301(5637), 1211–1216.

Song, A., Wu, D., Fan, Q., Tian, C., Chen, S., Guan, Z. et al. (2016) Transcriptome-wide identification and expression profiling analysis of chrysanthemum trihelix transcription factors. *International Journal of Molecular Sciences*, 17(2), 198.

Song, X., Hu, J., Wu, T., Yang, Q., Feng, X., Lin, H. et al. (2021) Comparative analysis of long noncoding RNAs in angiosperms and characterization of long noncoding RNAs in response to heat stress in Chinese cabbage. *Horticulture Research*, 8(1), 48.

Song, X., Li, N., Guo, Y., Bai, Y., Wu, T., Yu, T. et al. (2021) Comprehensive identification and characterization of simple sequence repeats based on the whole-genome sequences of 14 forest and fruit trees. *Forestry Research*, 1(1), 7–10.

Song, X., Li, Y. & Hou, X. (2013) Genome-wide analysis of the AP2/ERF transcription factor superfamily in Chinese cabbage (*Brassica rapa* ssp. pekinensis). *BMC Genomics*, 14(1), 573.

Song, X., Sun, P., Yuan, J., Gong, K., Li, N., Meng, F. et al. (2021) The celery genome sequence reveals sequential paleo-polyploidizations, karyotype evolution and resistance gene reduction in apiales. *Plant Biotechnology Journal*, 19(4), 731–744.

Song, X., Wang, J., Ma, X., Li, Y., Lei, T., Wang, L. et al. (2016) Origination, expansion, evolutionary trajectory, and expression bias of AP2/ERF superfamily in *Brassica napus*. *Frontiers in Plant Science*, 7, 1186.

Song, X., Wang, J.P., Sun, P.C., Ma, X., Yang, Q.H., Hu, J.J. et al. (2020) Preferential gene retention increases the robustness of cold regulation in Brassicaceae and other plants after polyploidization. *Horticulture Research*, 7, 20.

Song, X., Wei, Y., Xiao, D., Gong, K., Sun, P., Ren, Y. et al. (2021) *Brassica carinata* genome characterization clarifies U's triangle model of evolution and polyploidy in Brassica. *Plant Physiology*, 186(1), 388–406.

Taylor, J. & Raes, J. (2004) Duplication and divergence: the evolution of new genes and old ideas. *Annual Review of Genetics*, 38, 615–643.

Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D. et al. (2020) SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272.

Völz, R., Kim, S.K., Mi, J., Mariappan, K.G., Guo, X., Bigeard, J. et al. (2018) The trihelix transcription factor GT2-like 1 (GTL1) promotes salicylic acid metabolism, and regulates bacterial-triggered immunity. *PLoS Genetics*, 14(10), e1007708.

Wang, C., Wang, Y., Pan, Q., Chen, S., Feng, C., Hai, J. et al. (2019) Comparison of trihelix transcription factors between wheat and *Brachypodium distachyon* at genome-wide. *BMC Genomics*, 20(1), 142.

Wang, H., Chen, Y., Wu, X., Long, Z., Sun, C., Wang, H. et al. (2018) A potato STRUBBELIG-RECEPTOR FAMILY member, StLRPK1, associates with StSERK3A/BAK1 and activates immunity. *Journal of Experimental Botany*, 69(22), 5573–5586.

Wang, S., Li, L., Li, H., Sahu, S.K., Wang, H., Xu, Y. et al. (2020) Genomes of early-diverging streptophyte algae shed light on plant terrestrialization. *Nature Plants*, 6(2), 95–106.

Wang, T., Hu, J., Ma, X., Li, C., Yang, Q., Feng, S. et al. (2020) Identification, evolution and expression analyses of whole genome-wide TLP gene family in *Brassica napus*. *BMC Genomics*, 21(1), 264.

Wang, W., Wu, P., Liu, T., Ren, H., Li, Y. & Hou, X. (2017) Genome-wide analysis and expression divergence of the trihelix family in *Brassica rapa*: insight into the evolutionary patterns in plants. *Scientific Reports*, 7(1), 6463.

Wang, Y., Tang, H., Debarry, J.D., Tan, X., Li, J., Wang, X. et al. (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40(7), e49.

Wang, Z., Liu, Q., Wang, H., Zhang, H., Xu, X., Li, C. et al. (2016) Comprehensive analysis of trihelix genes and their expression under biotic and abiotic stresses in *Populus trichocarpa*. *Scientific Reports*, 6, 36274.

Wang, Z., Zhao, K., Pan, Y., Wang, J., Song, X., Ge, W. et al. (2018) Genomic, expressional, protein-protein interactional analysis of trihelix transcription factor genes in *Setaria italia* and inference of their evolutionary trajectory. *BMC Genomics*, 19(1), 665.

Win, K., Yamagata, Y., Doi, K., Uyama, K., Nagai, Y., Toda, Y. et al. (2017) A single base change explains the independent origin of and selection for the nonshattering gene in African rice domestication. *The New Phytologist*, 213(4), 1925–1935.

Xiao, J., Hu, R., Gu, T., Han, J., Qiu, D., Su, P. et al. (2019) Genome-wide identification and expression profiling of trihelix gene family under abiotic stresses in wheat. *BMC Genomics*, 20(1), 287.

Xu, H., Shi, X., He, L., Guo, Y., Zang, D., Li, H. et al. (2018) *Arabidopsis thaliana* trihelix transcription factor AST1 mediates salt and osmotic stress tolerance by binding to a novel AGAG-box and some GT motifs. *Plant & Cell Physiology*, 59(5), 946–965.

Yu, C., Song, L., Song, J., Ouyang, B., Guo, L., Shang, L. et al. (2018) ShCIGT, a trihelix family gene, mediates cold and drought tolerance by interacting with SnRK1 in tomato. *Plant Science*, 270, 140–149.

Yu, T., Bai, Y., Liu, Z., Wang, Z., Yang, Q., Wu, T. et al. (2022) Large-scale analyses of heat shock transcription factors and database construction based on whole-genome genes in horticultural and representative plants. *Horticulture Research*, 9, uhac035.

Yu, T., Ma, X., Liu, Z., Feng, X., Wang, Z., Ren, J. et al. (2022) TVIR: a comprehensive vegetable information resource database for comparative and functional genomic studies. *Horticulture Research*, 9, uhac213.

Zhang, Y., Zhang, Y., Li, B., Tan, X., Zhu, C., Wu, T. et al. (2022) Polyploidy events shaped the expansion of transcription factors in Cucurbitaceae and exploitation of genes for tendril development. *Horticultural Plant Journal*, 8(5), 562–574.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Wu, T., Yang, Q., Zhou, R., Yu, T., Shen, S., Cao, R. et al. (2023) Large-scale analysis of trihelix transcription factors reveals their expansion and evolutionary footprint in plants. *Physiologia Plantarum*, 175(5), e14039. Available from: https://doi.org/10.1111/ppl.14039