

Genome-wide Identification and Expression Analysis of the 14-3-3 Gene Family in Cotton (*Gossypium Hirsutum* L.)

wenju gao (✉ gaowenju120@gmail.com)

Xinjiang Agricultural University

qin chen

Xinjiang Agricultural University

jielyn zhao

Xinjiang Agricultural University

peng wang

Xinjiang Agricultural University

duolu li

Xinjiang Agricultural University

kai zheng

Xinjiang Agricultural University

yilei long

Xinjiang Agricultural University

yang jiao

Xinjiang Agricultural University

yuxiang wang

Xinjiang Agricultural University

shiwei geng

Xinjiang Agricultural University

xuening su

Xinjiang Agricultural University

yanying qu

Xinjiang Agricultural University

quanjia chen

Xinjiang Agricultural University

Research Article

Keywords: Cotton, 14-3-3 protein, Evolution analysis, Functional expression analysis

Posted Date: January 7th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-135544/v1>

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

[Read Full License](#)

Genome-wide identification and expression analysis of the 14-3-3 gene family in cotton (*Gossypium hirsutum* L.)

Wen-Ju Gao, Qin Chen, Jie-Yin Zhao, Peng Wang, Duo-Lu Li, Kai Zheng, Yi-Lei Long, Yang Jiao, Yu-Xiang Wang, Shi-Wei Geng, Xue-Ning Su, Yan-Ying Qu & Quan-Jia Chen

Abstract

Background: As a ubiquitous acid-regulating protein family in eukaryotes, 14-3-3 proteins are widely involved in the growth and development of plants. With the development of the third-generation sequencing technology and the smooth completion of the cotton genome work, it is possible to explore the existence and distribution of the 14-3-3 protein family in cotton.

Results: In this paper, 33, 33, 17 and 18 members were identified from this family in *Gossypium hirsutum*((AD)₁), *G. barbadense*((AD)₂), *G. arboreum*(A₂) and *G. raimondii*(D₅), respectively. In particular, evolution analysis, structure analysis and functional expression analysis of this protein family in *G. hirsutum*((AD)₁) were carried out. The results showed that compared with *Arabidopsis* and rice, the phylogenetic tree and gene structure clearly divided the 14-3-3 protein family into two subgroups in *G. hirsutum*((AD)₁), the ϵ group and the non- ϵ group; Analysis of transcriptome expression patterns revealed that this family was significantly induced to express under abiotic stress; Most 14-3-3 proteins have a large number of cis-acting elements related to growth, development and abiotic stress in the promoter region, among which elements related to drought stress account for the largest proportion; The results of qRT-PCR showed that the expression of 14-3-3 protein had significant differences under drought stress.

Conclusions: In summary, this study signified that the 14-3-3 protein family is relatively conserved in the evolutionary expansion of cotton, and may be involved in the growth and development of plants and the mechanism of stress resistance. These results provide an important theoretical and experimental basis for further analysis and verification of the function of 14-3-3 protein in cotton.

Keywords: Cotton, 14-3-3 protein, Evolution analysis, Functional expression analysis

Background

14-3-3 protein is the main soluble protein found in bovine brain tissue by Moore and Perez^[1] in 1967. In plants, it was first found in *Arabidopsis*^[2] and other plants. Because this protein is often part of the complex that binds to G-box when it is discovered, it is also called "G-box factor 14-3-3", or "GF14" for short^[3]. According to the gene

structure, the GF14 family in rice^[4], *Arabidopsis*^[5], *alfalfa*^[6], *soybean*^[7], *grape*^[8], *poplar*^[9] and other species can be roughly divided into two subgroups: ϵ group and non- ϵ group^[10]. The ϵ group generally has 6-7 exons and 4-6 introns, and the non- ϵ group generally has 4 exons and 3 introns^[11,12].

In plants, the main function of GF14 protein is to act as a key regulator of primary metabolism and signal transduction, and participate in the regulation of plant growth and development and response to adversity stresses. In maize, Jahn T^[13] et al. found that GF14 protein is a positive regulator of plasma membrane H⁺-ATPase enzyme activity, involved in regulating osmotic pressure, and affecting water transport in plants; In tobacco, Yohsuke Takahashi et al.^[14,15] proved that GF14 protein can regulate the distribution of endogenous gibberellins, thereby indirectly regulating plant stem elongation, seed germination and flowering processes; In apple and wheat^[16,17], GF14 protein reduces cell oxidative damage by regulating the expression of stress-related genes to clear reactive oxygen species in signal pathways, which can be considered as a positive regulator for salt and drought stress responses; In *Arabidopsis*^[18], overexpression of GF14 protein can enhance the cold resistance and adversity stresses response; In cotton^[19], Juqiang Yan et al. overexpressed AtGF14 λ , a subtype gene encoding GF14 protein in *Arabidopsis*, in cotton. Compared with wild-type cotton plants, the overexpression lines showed the characteristics of long-lasting greenness and drought resistance. In short, these studies have shown that GF14 protein plays an important role in plant stress resistance, growth and development.

Cotton is an important economic crop in the world. In recent years, as the global climate changes, its planting production is often affected by various environmental pressures. At present, the research of cotton GF14 protein in stress resistance has not been reported, so it is of great significance to study the GF14 protein family in cotton. Based on the cotton genome database^[20,21], this study systematically identified members of the GF14 protein family, focusing on the evolution and transmission of the GF14 protein family in cotton and its response function under abiotic stress. It lays a good foundation for further analysis of the molecular mechanism of GF14 protein family in response to various adversities during the growth and development of cotton.

Results

Identification of Cotton GF14s and Analysis of physicochemical property

Through the local Blast database and NCBI-CDD database^[30], the cotton GF14 protein family were screened and verified. Finally, 33, 33, 17 and 18 Members of this family were identified in four cotton species, including *Gossypium hirsutum*((AD)₁), *G. barbadense*((AD)₂), *G. arboreum*(A₂) and *G. raimondii*(D₅), respectively (Fig. S1). And the family members named according to the sequence of the gene position on the chromosome(Table 1). CottonGen^[34] database comparison found(Table 1), 33 GhGF14s CDS length is 285bp(GhGF14-32) - 810bp(GhGF14-02, 18), protein sequence length is 94aa(GhGF14-32) - 269aa(GhGF14-02, 18); ExPASy^[31] predicted

that the molecular weight of 33 GhGF14s was 10.80kDa(GhGF14-32) - 30.40kDa(GhGF14-18), and the isoelectric point was 4.69(GhGF14-05, 13) - 6.29(GhGF14-32); The prediction of subcellular localization of Softberry ProtComp9.0 shows that 30 of the 33 GhGF14s are located in the cytoplasm and nucleus, and 3 are located in the nucleus and cell membrane (GhGF14-07, 19, 23). It can be seen that there are big differences in the physical and chemical properties of GhGF14s, indicating that GhGF14s may have a more complicated division of labor and functions during the development of cotton.

Gene_ID	Gene name	Chromosome location	CDS Length (bp)	Pep Length (aa)	PI	Molecular weight (kDa)	sub-cellular location	Group
GH_A01G0120	GhGF14-01	A01:907655-909600-	786	261	4.79	29.33	Cytoplasm_and_Nucleus	non-ε
GH_A01G1276	GhGF14-02	A01:39234674-39237265-	810	269	4.84	30.39	Cytoplasm_and_Nucleus	non-ε
GH_A02G0841	GhGF14-03	A02:16307796-16309959-	759	252	4.74	28.53	Cytoplasm_and_Nucleus	non-ε
GH_A03G1993	GhGF14-04	A03:107387974-107389708+	786	261	4.74	29.26	Cytoplasm_and_Nucleus	non-ε
GH_A04G1376	GhGF14-05	A04:83478393-83480226+	786	261	4.69	29.22	Cytoplasm_and_Nucleus	non-ε
GH_A05G1561	GhGF14-06	A05:14612017-14613882-	765	254	4.85	28.96	Cytoplasm_and_Nucleus	ε
GH_A05G2209	GhGF14-07	A05:21373609-21375687+	759	252	4.71	28.50	Nucleus_and_Membrane	non-ε
GH_A05G2242	GhGF14-08	A05:21685232-21686305-	807	268	4.72	30.24	Cytoplasm_and_Nucleus	non-ε
GH_A05G2260	GhGF14-09	A05:21904939-21907149+	762	253	4.80	28.85	Cytoplasm_and_Nucleus	ε
GH_A05G2680	GhGF14-10	A05:28818674-28819734-	801	266	4.90	30.13	Cytoplasm_and_Nucleus	non-ε
GH_A06G0365	GhGF14-11	A06:4326636-4327669-	777	258	4.70	29.03	Cytoplasm_and_Nucleus	non-ε
GH_A06G0380	GhGF14-12	A06:4499939-4501821+	756	251	4.76	28.64	Cytoplasm_and_Nucleus	ε
GH_A07G0784	GhGF14-13	A07:9610675-9612667+	789	262	4.69	29.60	Cytoplasm_and_Nucleus	non-ε
GH_A07G1832	GhGF14-14	A07:57732873-57734797+	789	262	4.70	29.42	Cytoplasm_and_Nucleus	non-ε
GH_A13G0728	GhGF14-15	A13:15421114-15423105+	786	261	4.78	29.51	Cytoplasm_and_Nucleus	non-ε
GH_A13G0888	GhGF14-16	A13:31329265-31331175-	711	236	5.16	26.64	Cytoplasm_and_Nucleus	ε
GH_D01G0111	GhGF14-17	D01:825496-827457-	786	261	4.81	29.33	Cytoplasm_and_Nucleus	non-ε
GH_D01G1332	GhGF14-18	D01:24063415-24065943+	810	269	4.79	30.40	Cytoplasm_and_Nucleus	non-ε
GH_D02G0861	GhGF14-19	D02:14190589-14192762-	759	252	4.73	28.44	Nucleus_and_Membrane	non-ε

GH_D02G2162	GhGF14-20	D02:66107615-66109355+	786	261	4.73	29.25	Cytoplasm_and_Nucleus	non-ε
GH_D04G1722	GhGF14-21	D04:52939506-52941344+	786	261	4.74	29.23	Cytoplasm_and_Nucleus	non-ε
GH_D05G1589	GhGF14-22	D05:13421094-13422924-	777	258	4.90	29.25	Cytoplasm_and_Nucleus	ε
GH_D05G2234	GhGF14-23	D05:19488600-19490677+	759	252	4.71	28.50	Nucleus_and_Membrane	non-ε
GH_D05G2268	GhGF14-24	D05:19795734-19796811-	807	268	4.72	30.24	Cytoplasm_and_Nucleus	non-ε
GH_D05G2282	GhGF14-25	D05:19945167-19947370+	762	253	4.80	28.88	Cytoplasm_and_Nucleus	ε
GH_D05G2698	GhGF14-26	D05:25186907-25193203-	672	223	4.91	25.61	Cytoplasm_and_Nucleus	non-ε
GH_D06G0355	GhGF14-27	D06:4187799-4189677+	756	251	4.75	28.68	Cytoplasm_and_Nucleus	ε
GH_D07G0785	GhGF14-28	D07:8517104-8519264+	789	262	4.76	29.60	Cytoplasm_and_Nucleus	non-ε
GH_D07G1863	GhGF14-29	D07:37722088-37723873+	792	263	4.81	29.53	Cytoplasm_and_Nucleus	non-ε
GH_D07G1925	GhGF14-30	D07:44363054-44365105-	789	262	4.77	29.44	Cytoplasm_and_Nucleus	non-ε
GH_D13G0539	GhGF14-31	D13:6580573-6582577-	786	261	4.89	29.44	Cytoplasm_and_Nucleus	non-ε
GH_D13G0612	GhGF14-32	D13:8021308-8021592-	285	94	6.29	10.80	Cytoplasm_and_Nucleus	non-ε
GH_D13G0845	GhGF14-33	D13:15329160-15330847+	720	239	4.96	26.94	Cytoplasm_and_Nucleus	ε

Table 1. Analysis of the members of the GF14 protein family and related physical and chemical properties of upland cotton.

Chromosome Location and Phylogenetic Tree Analysis

In order to explore the homology of GF14s in *G. hirsutum*((AD)₁), *G. barbadense*((AD)₂), *G. arboreum*(A₂) and *G. raimondii*(D₅), we first locate GF14s on the chromosome (Fig. 1BCDE). In *G. hirsutum*((AD)₁) and *G. barbadense*((AD)₂), the same number of GF14s are distributed on the same 15 chromosomes (Fig. 1BC), in particular, in their respective AD subgroup chromosomes, among them, there are 2 GF14s on A13 chromosome and 3 GF14s on D13 chromosome; In *G. arboreum*(A₂), the distribution of GF14s is roughly the same as the distribution of GF14s on the chromosomes of *G. hirsutum*((AD)₁) and *G. barbadense*((AD)₂) A subgroup (Fig. 1D), but it is worth noting that there are no GF14s on the Chr02 chromosome of *G. arboreum*(A₂), and two on the Chr03 chromosome, And *G. hirsutum*((AD)₁) and *G. barbadense*((AD)₂) each have one GF14s on A02 and A03 chromosomes; In *G. raimondii*(D₅), the distribution and number of GF14s on the chromosomes are somewhat different from the distribution on the chromosomes of *G. hirsutum*((AD)₁) and *G. barbadense*((AD)₂) D subgroup. This indicates that part of the gene may be lost and transferred during the evolution of the D subgroup in cotton.

A phylogenetic tree was constructed from a total of 101 protein sequences of the GF14 protein family in the four cotton species(Table S1), which can obviously be divided into

4 subgroups (Fig. 1A). The phylogenetic tree shows that the GF14s exist in the same position on the chromosomes in *G. hirsutum*((AD)₁) and *G. barbadense*((AD)₂) are tightly clustered, while *G. arboreum*(A₂) and *G. raimondii*(D₅) also have GF14s on the chromosomes of the AD subgroups that are closest to their respective relationships Gathered in a subgroup. Surprisingly, although divided into four major subgroups, the number of genes in *G. hirsutum*((AD)₁) and *G. barbadense*((AD)₂) in each subgroup is basically about twice that of *G. arboreum*(A₂) and *G. raimondii*(D₅). It is in line with the evolutionary relationship of cotton species.

Based on the study of GF14 protein family in *Arabidopsis*^[5] and rice^[4], phylogenetic trees belonging to *G. hirsutum*((AD)₁), *Arabidopsis* and rice were constructed (Fig. 2). Similarly, the phylogenetic tree is divided into four subgroups. Based on the clustering position of AtGF14, the four subgroups are classified into two broadgroups unique to the GF14 family: ϵ group and non- ϵ group.

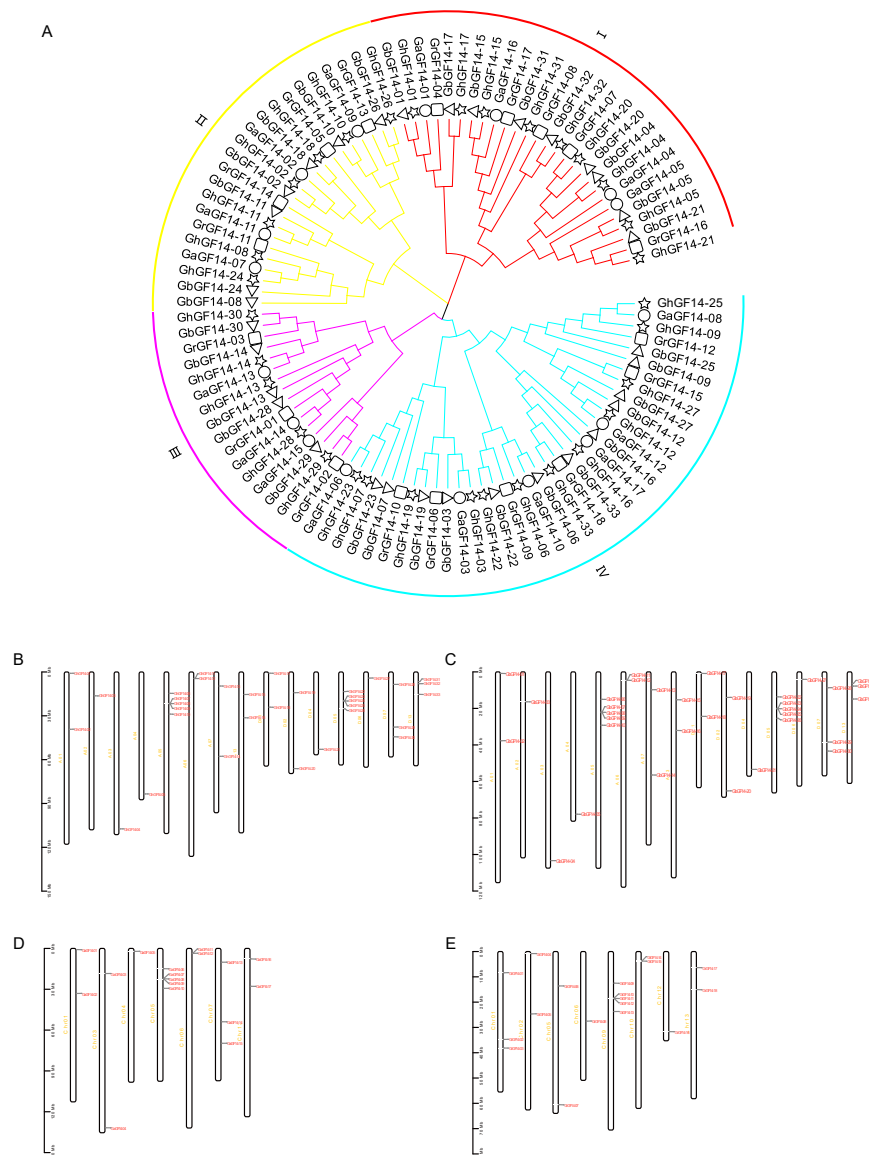


Fig. 1. Chromosome location and phylogenetic tree of cotton GF14 protein family. (A) NJ evolutionary tree constructed with 101 GF14 protein sequences of *Gossypium hirsutum*((AD)₁), *G. barbadense*((AD)₂), *G. arboreum*(A₂) and *G. raimondii*(D₅). The stars represent *G. hirsutum*((AD)₁), triangles represent *G. barbadense*((AD)₂), circles represent *G. arboreum*(A₂), and squares represent *G. raimondii*(D₅). (B) Chromosome location of GF14 gene in *G. hirsutum*((AD)₁). (C) Chromosome location of GF14 gene in *G. barbadense*((AD)₂). (D) Chromosome location of GF14 gene in *G. arboreum*(A₂). (E) Chromosome location of GF14 gene in *G. raimondii*(D₅).

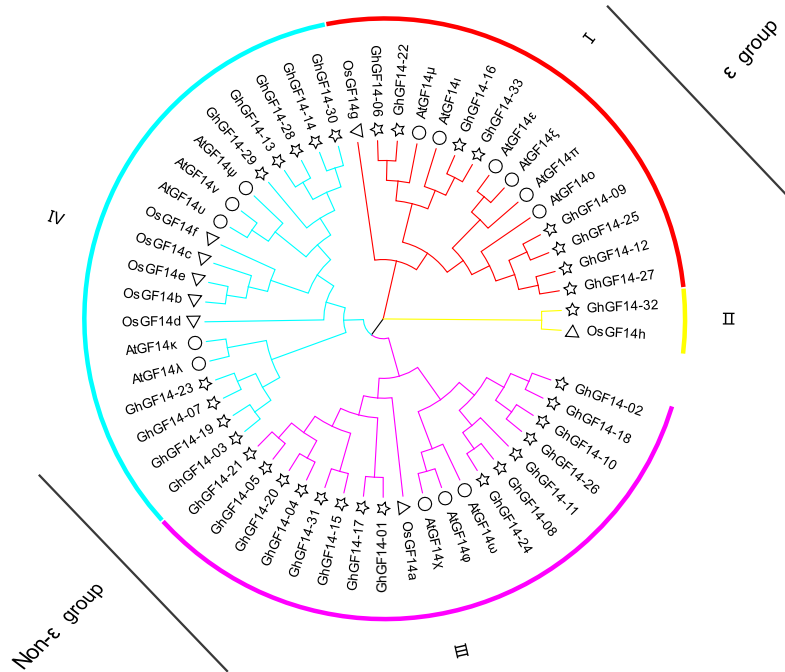


Fig. 2. Phylogenetic tree of GF14 protein gene family in *Gossypium hirsutum*((AD)₁), *Arabidopsis* and rice. The stars represent GF14s in *G. hirsutum*((AD)₁). The circle represents GF14s in *Arabidopsis*. The triangles represent GF14s in rice. The names of GF14s in *Arabidopsis* and rice are derived from previous studies.

Identification and analysis of GhGF14 gene homology

Compared with other species, *Gossypium hirsutum*((AD)₁) is an allotetraploid (AADD), and studying its own homology has great significance for the evolution of GF14 protein family in cotton. The homology analysis of GhGF14s on the chromosomes of the *G. hirsutum*((AD)₁) AD subgroup found (Fig. 3A) that the number and distribution of GhGF14s on the chromosomes of group A and group D are similar. This shows that the evolutionary expansion of this family is somewhat conservative from the perspective of cotton seed evolution; GhGF14s has 54 pairs of non-tandem repeats, 2 pairs of tandem repeats, that is, 56 pairs of paralogous genes, of which there are 10 pairs between A and A, 34 between A and D, and D and D there are 12 pairs; Most genes have one-to-many collinearity, which proves that these genes may play an important role in the evolution and expansion of this family.

Rice and *Arabidopsis* are model plants in monocotyledonous(dicotyledonous) plants, which play a role as species representative to some extent ^[40,41]. Using the collinearity analysis among the genomes of *G. hirsutum*((AD)₁), rice and *Arabidopsis thaliana*, we can understand the evolutionary relationship of the GhGF14 protein family in monocots and dicots. As shown in Fig. 3B, there are 19 pairs of orthologous genes between the

genomes of *Arabidopsis* and *G. hirsutum*((AD)₁), indicating that they have a relatively close evolutionary relationship; However, the collinearity of this family between rice and *G. hirsutum*((AD)₁) is zero, which indicates that the members of this family may have evolved and expanded mainly after the angiosperms split into monocots and dicots. The results of this analysis provide important clues for further research on the function of the protein in *G. hirsutum*((AD)₁) using the existing research results of the protein in *Arabidopsis*.

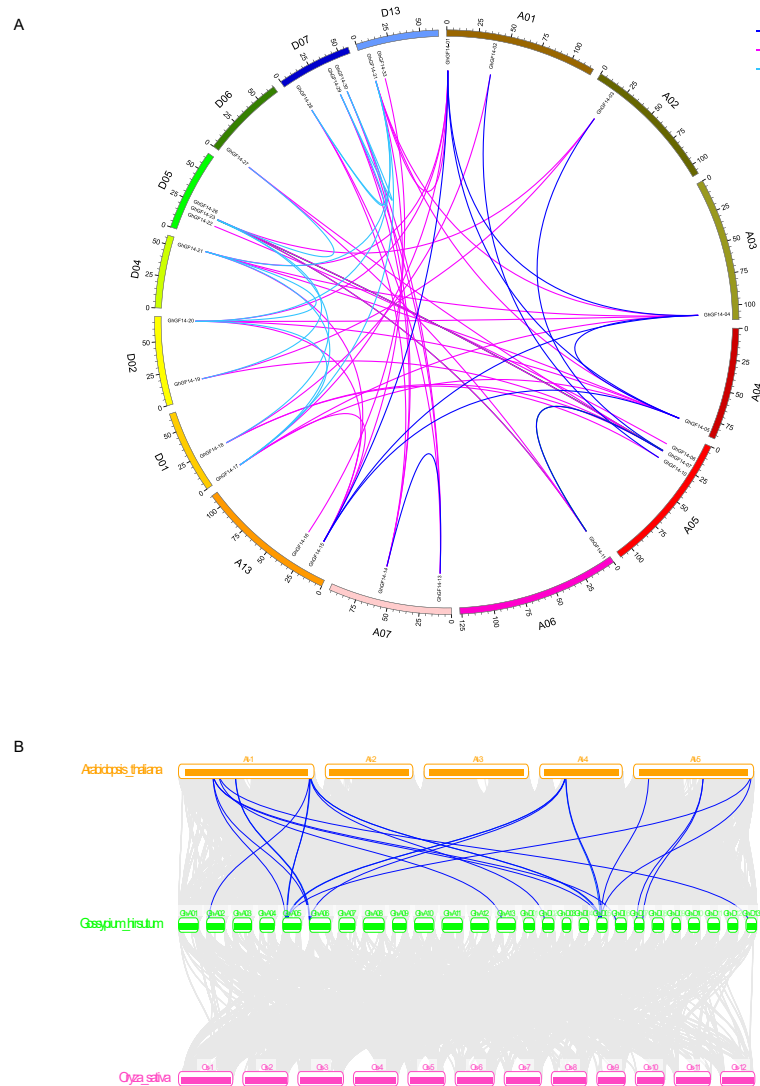


Fig. 3. Homology identification and analysis of GhGF14s gene. **(A)** The homology of GhGF14s between the chromosomes in the AD group of *G. hirsutum*((AD)₁). **(B)** The homology of GhGF14s among the genomes of *G. hirsutum*((AD)₁), rice and *Arabidopsis*.

Gene structure analysis of GhGF14 protein family

The GF14 protein family has its own unique gene structure^[10]. Using the evolutionary tree (Fig. 2) for clustering, the conservative motif predicted by MEME^[38] and the GhGF14s gene structure information obtained from CottonGen^[34]. The cotton genome database, were combined and spliced to obtain the unique GF14 proteins family Grouping mode (Fig. 4). Through the clustering of phylogenetic tree, genes with high

sequence similarity and similar gene structure are clustered together (Fig. 4A).

According to the structural information of the gene (Fig. 4C), the ϵ group of this family has 6-7 exons and 4-6 introns, Non- ϵ group has 4 exons and 3 introns, and the first exon of the non- ϵ group is longest. This is consistent with the ϵ group and non- ϵ group classification criteria specific to the GF14 protein family^[11,12]. And from this, the exact ϵ group and non- ϵ group of the GhGF14 protein family were obtained (Table 1): The ϵ group has 9 genes (GhGF14-06, 09, 12, 16, 22, 25, 27, 33), and the non- ϵ group has 24 genes (GhGF14-01, 02, 03, 04, 05, 07, 08, 10, 11, 13, 14, 15, 17, 18, 19, 20, 21, 23, 24, 26, 28, 29, 30, 31). The relatively short GhGF14 gene (GhGF14-32) at the bottom may be an imperfect annotation information in the genome file, so it will not be classified for the time being. The composition of conservative motifs, in terms of number and location (Fig. 4B), is roughly similar in same cluster subgroup, indicating that these family members may have similar functions and regulatory mechanisms; In addition to the same motif1, 2, 3, 4, 5, 6, 7, motif8 unique to the non- ϵ group and motif9 unique to the ϵ group deserve our attention, because this may be a grouping exercise of the GF14 protein family. One of the basis for different functions.

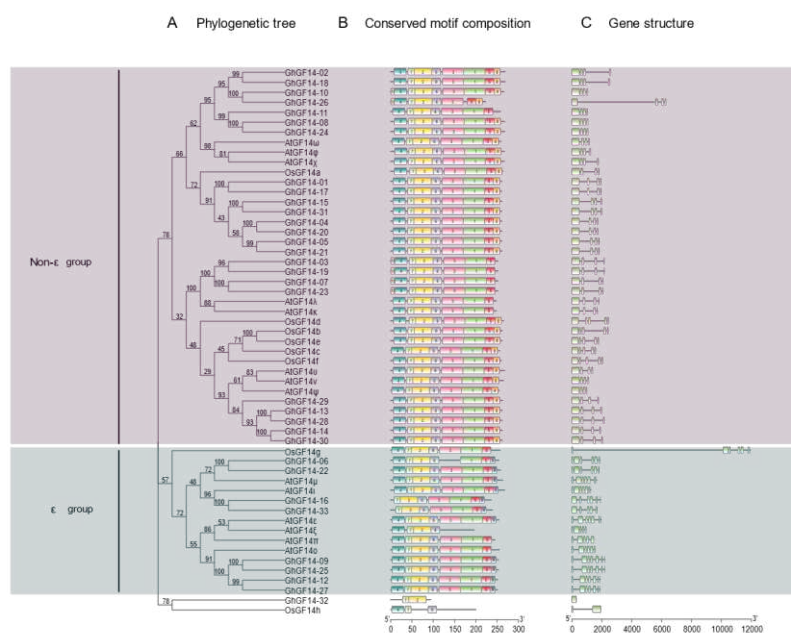


Fig. 4. Gene structure diagram of GF14 protein family in *G. hirsutum*((AD)₁), rice and *Arabidopsis*. (A) Phylogenetic tree of GF14s protein family in *G. hirsutum*((AD)₁), rice and *Arabidopsis*. (B) Conserved motif structure prediction. (C) The distribution of exons and introns in the CDS region of GF14s.

Analysis of the expression pattern of GhGF14s gene family

The expression pattern of genes determines the function of genes to a certain extent. The analysis results of tissue-specific expression patterns (Fig. 5) show that the expression patterns of GhGF14 gene family in tissues are quite different. Among them, GhGF14-09, 25, 05, and 21 have high expression levels in different tissues, indicating that it may participate in many important links in the growth and development of cotton; GhGF14-16 and 33 are only highly expressed in stamens, indicating that they may be

involved in the regulation of cotton pollen development; GhGF14-11 and 08 are highly expressed in petals and pistils, indicating that they may play a role in the reproduction and growth of cotton; GhGF14-13, 28, 14, 30, 20, 04, 23, etc. are highly expressed in roots, stems and leaves, indicating that it may play an important role in the process of cotton vegetative growth.

Similarly, under different abiotic stresses, the expression patterns of GhGF14 family members also showed great differences (Fig. 6). With the increasing time of different abiotic stresses, GhGF14s roughly presents three expression patterns: high expression, low expression and micro-expression or no expression. In particular, under cold stress, 11 GhGF14 genes showed high expression (GhGF14-09, 25, 05, 21, 01, 14, 23, 04, 20, 30, 17), and increased with the stress time, the amount of expression generally shows a continuous downward trend; Under heat stress, 12 GhGF14 genes showed obviously high expression (GhGF14-09, 25, 05, 21, 01, 14, 23, 04, 20, 30, 12, 27), and with the increase of stress time, the expression level generally shows a continuous upward trend, which is the opposite to cold stress; Under drought stress, 13 GhGF14 genes showed differentially high expression (GhGF14-09, 25, 05, 21, 01, 14, 23, 04, 20, 30, 12, 27, 24), and with the time of stress increase, the amount of expression presents a trend of upward volatility; Under salt stress, 13 GhGF14 genes showed differentially high expression (GhGF14-09, 25, 05, 21, 01, 14, 23, 04, 20, 30, 12, 27, 17), and with the time of stress growth, the amount of expression also shows a trend of upward volatility.

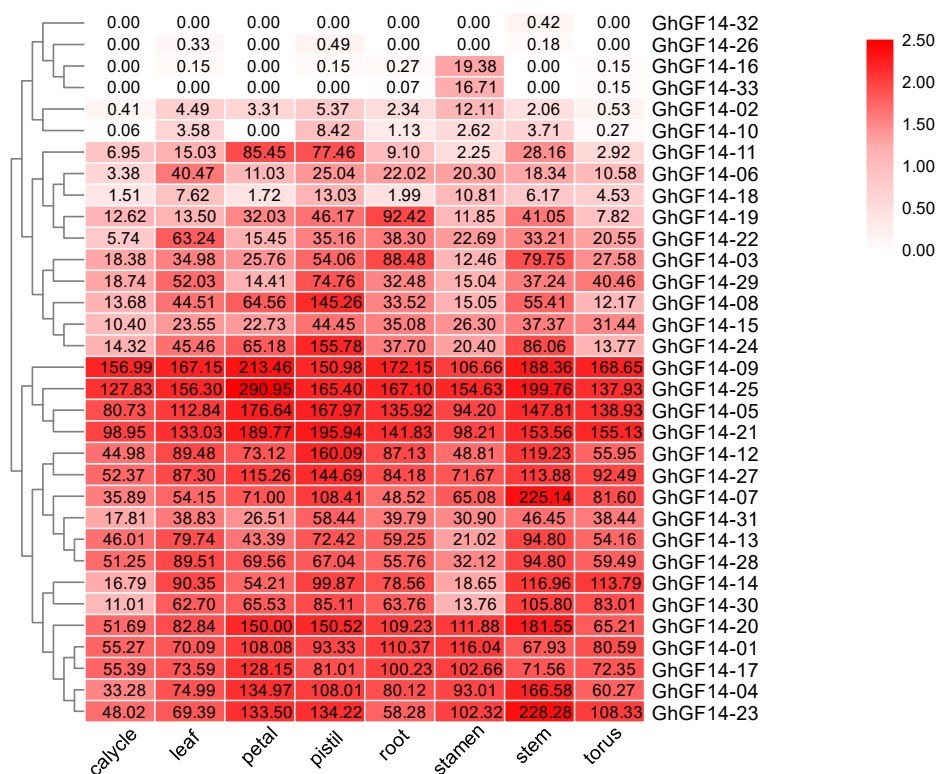


Fig. 5. Tissue specificity at the transcriptome level of the GhGF14 gene family. Use $\log_{10}(x+1)$ standardization for data processing, x is the original expression value, and the original expression value is shown in the figure; the red to white gradient represents the change of expression from high to low.

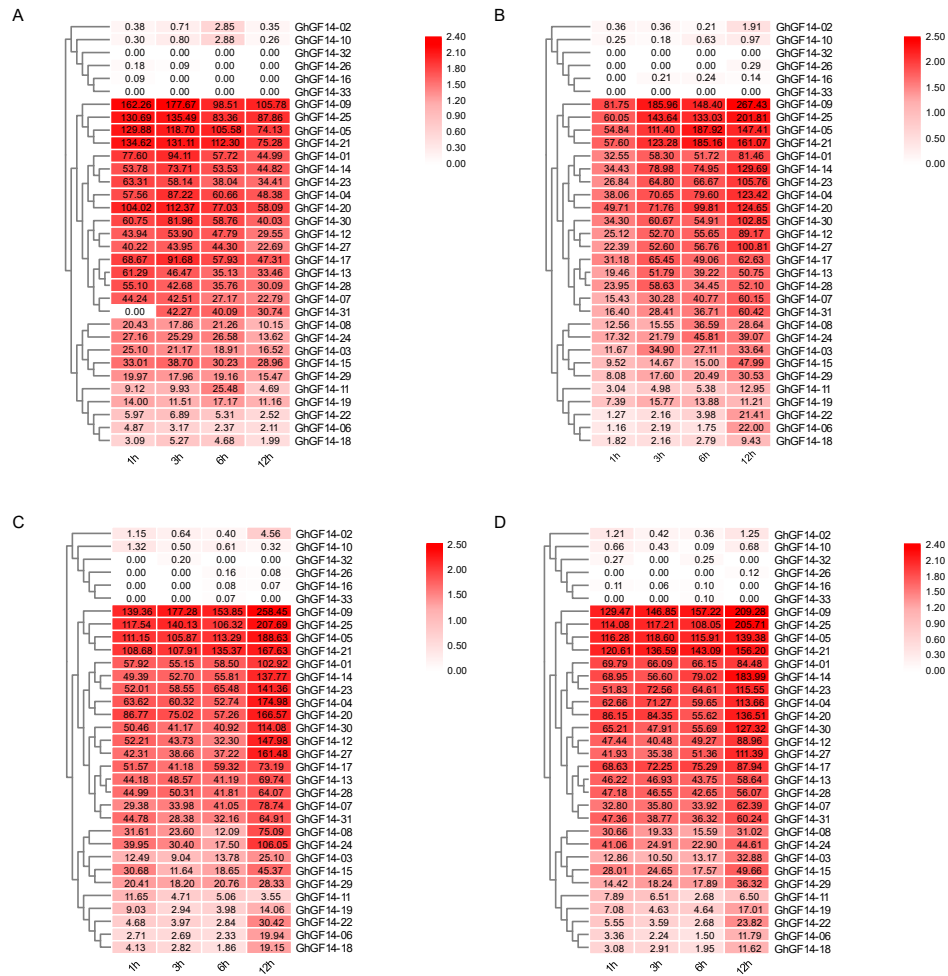


Fig. 6. Analysis of the expression pattern of GhGF14s gene family. **(A)** Changes in GhGF14s gene expression during 1h, 3h, 6h, 12h periods under cold stress; **(B)** Changes in GhGF14s gene expression during 1h, 3h, 6h, 12h periods under heat stress; **(C)** Changes in GhGF14s gene expression during 1h, 3h, 6h and 12h periods under drought stress; **(D)** Changes in GhGF14s gene expression during 1h, 3h, 6h and 12h periods under salt stress. Use $\log_{10}(x+1)$ standardization for data processing, x is the original expression value, and the original expression value is shown in the figure; the red to white gradient represents the change of expression from high to low.

Analysis of cis-acting elements in the promoter region

As a specific binding site involved in the initiation and regulation of protein transcription, the study of cis-acting elements in the gene promoter region is of great significance for understanding the overall regulation of plant gene expression^[42]. It can be seen from the figure (Fig. 5A) that there are multiple cis-acting elements related to stress resistance or growth and development in each promoter region of the GhGF14 protein family. Such as abscisic acid, auxin, salicylic acid, jasmonic acid, flavonoids and other hormone-related cis-acting elements, and low temperature, drought, defense and stress response and other anti-stress related cis-acting elements. This indicates that the GhGF14 protein family may be involved in different regulatory mechanisms for resistance to stresses during growth and development in plant.

The upsetplot (Fig. 5B) was constructed to reflect the distribution and proportion of these abiotic stress-related elements in the GhGF14s family. MYB(C/TAAACNA/G), studies have shown that it is a cis-acting element involved in drought, salt, and low temperature stress^[43], which is present in 100%(33/33) of the GhGF14s gene promoter region; MYC(CANNTG), which is considered to be an element involved in drought stress and ABA biosynthesis^[44], is present in 97%(32/33) of the GhGF14s gene promoter region; ABRE(ACGT), as an important element involved in ABA biosynthesis and drought stress^[45], is present in 64%(21/33) of the GhGF14s gene promoter region; W-box(TTGACC), which can be combined with WRKY transcription factor, participates in the response of plants to diseases, drought, ABA and other adversities^[46]. It exists in 55%(18/33) of the GhGF14s gene promoter region; TC rich repeats(GTTTTCTTAC), involved in plant defense and stress response^[47], exists in 30.3%(10/33) of the GhGF14s gene promoter region; DRE(CCGAC), a cis-acting element involved in drought, salt, and low temperature stress^[48], is present in 21.2%(7/33) of the GhGF14s gene promoter region; LTR(CCG AAA), a cis-acting element involved in low temperature stress^[49], is present in 33.3%(11/33) of the GhGF14s gene promoter region.

It can be seen that the distribution and proportion of cis-acting elements involved in stress resistance-related summation in the promoter region of GhGF14 protein family members are large. Especially drought-related cis-acting elements (MYB, MYC, ABRE, W-box, DRE, etc.) exist in almost every GhGF14 gene promoter region. This indicates that the GhGF14 protein family may play an important role in the process of plant for stress resistance. Therefore, it should be of great significance to study the GhGF14 protein family in the process of stress resistance in plant, especially drought-related response mechanisms.

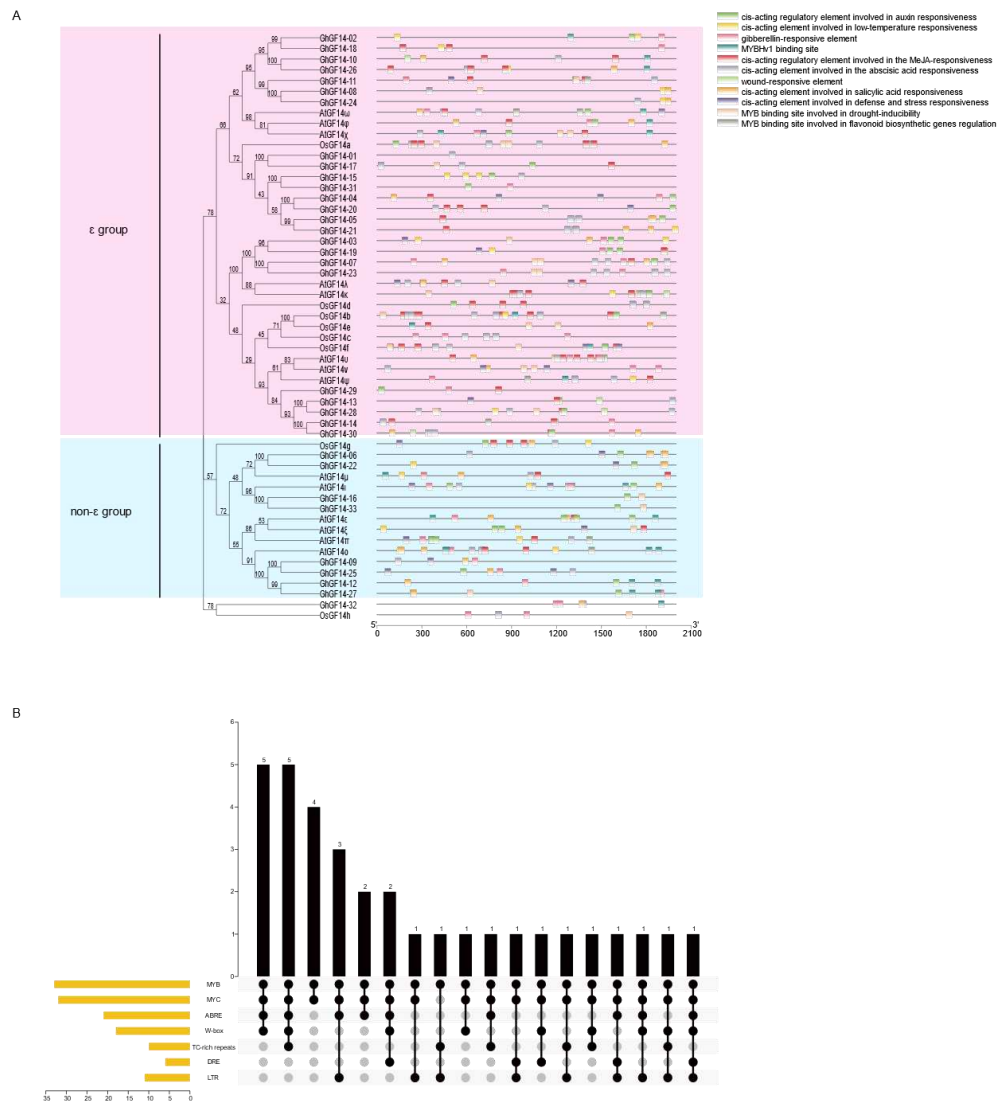


Fig. 7. Analysis of cis-acting elements in the GhGF14s promoter region. **(A)** The presentation of cis-acting elements related to stress resistance and growth and development in the GhGF14s promoter region based on the comparison between rice and Arabidopsis. **(B)** Upsetplot constructed by cis-acting elements related to stress resistance.

qRT-PCR analysis

According to the results of transcriptome expression analysis and the Upsetplot of cis-acting elements related to the promoter region (Fig. 5B), the GhGF14s gene with obvious expression changes under drought stress and the most intersection of drought-related cis-acting elements was specifically selected. Finally, seven GhGF14 genes were selected (GhGF14-09, 14, 20, 21, 23, 25, 30). At the same time, in order to verify the accuracy of RNA-seq expression analysis, the same leaf parts were selected for qRT-PCR analysis.

The results show (Fig. 7) the relative expression levels of 7 GhGF14s genes in the leaves of drought-resistant materials (KK1543) and drought-sensitive materials

(Xinluzao 26) in 1h, 3h and 6h under drought stress shows the opposite trend of change. And in the drought sensitive material (Xinluzao 26), there is a significant high expression compare to the normal treatment at some specific time periods, such as 6h, 24h, etc. This is apparently the same as the result of RNA-seq expression analysis.

In addition, in order to explore the specific expression of the GhGF14s protein family in different cotton tissues, the roots, stems and leaves of Xinluzao 26 at 0h and 6h were selected for qRT-PCR analysis under drought stress (Fig. 8). The results showed that 7 GhGF14 genes generally had higher expression levels in roots under normal 0h treatment, followed by stems, and compared with the expression levels of roots and stems under 6h stress treatment, they generally showed a significant downward trend; On the contrary, in the leaves, the expression level of the leaves under 0h normal treatment and 6h stress treatment showed an up-regulation trend, and there was a significant up-regulation in GhGF14-21 and GhGF14-30.

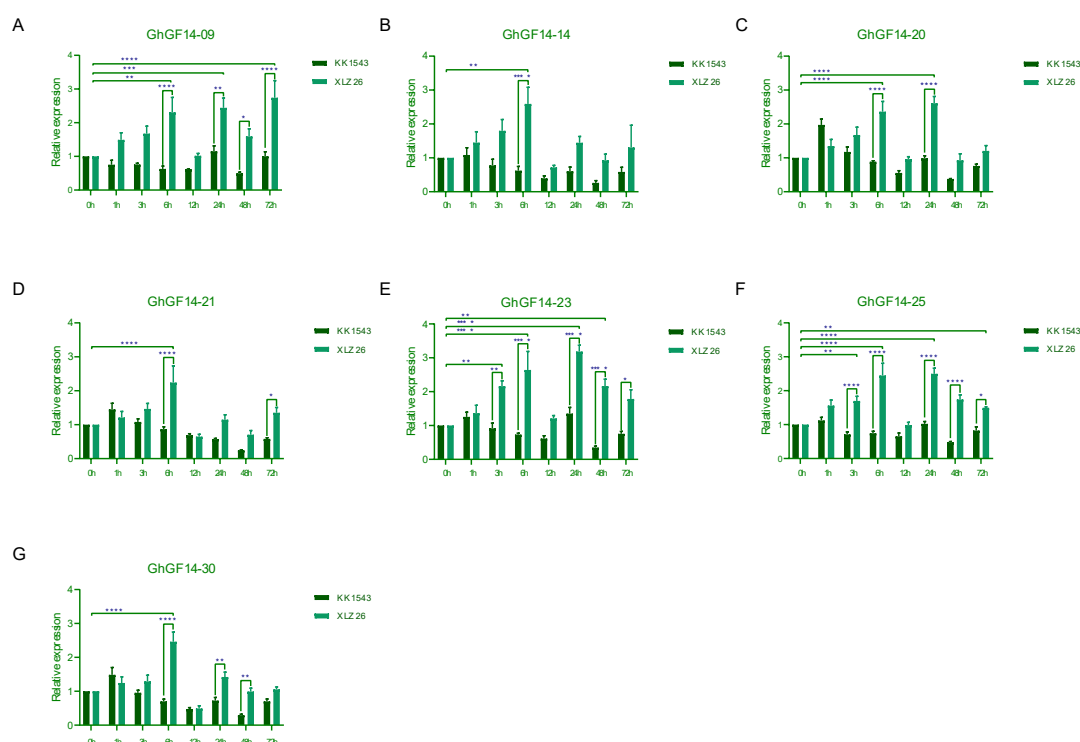


Fig. 8. Drought stress qRT-PCR analysis. **ABCDEFG** respectively represents the corresponding gene name. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

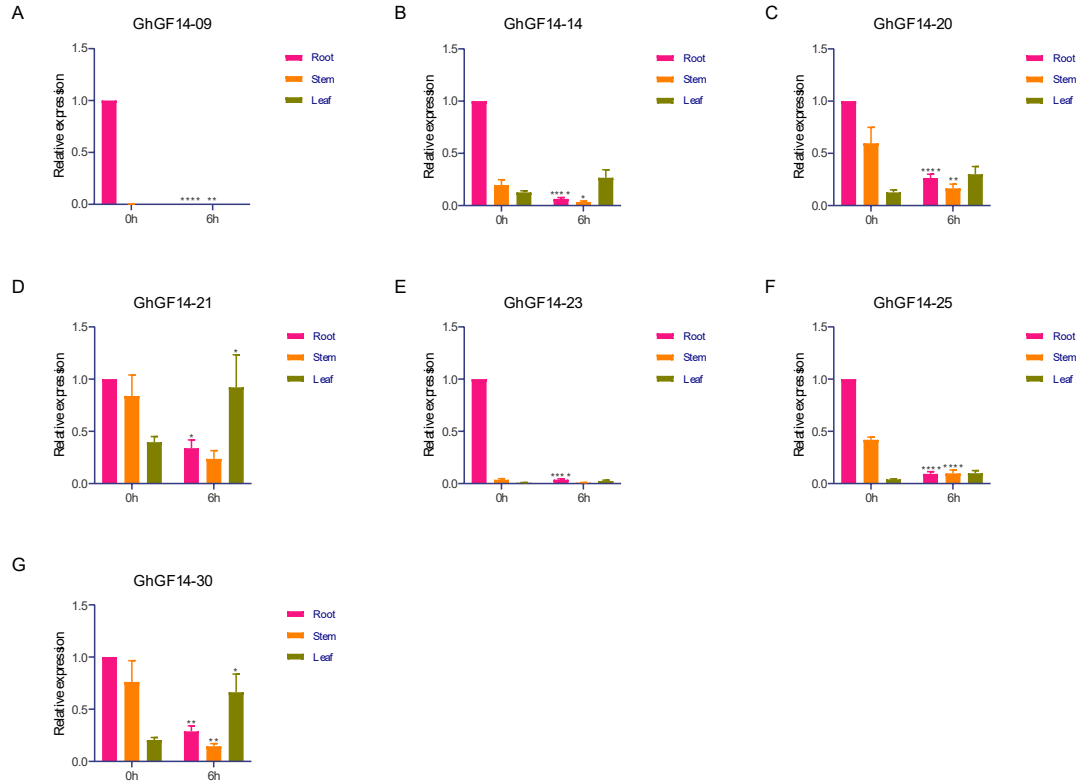


Fig. 9. QTR-PCR tissue specificity analysis. ABCDEFG respectively represents the corresponding gene name. * P<0.05, ** P<0.01, *** P<0.001, **** P<0.0001.

Discussion

GF14 protein is highly conserved in eukaryotes, and it can form homodimers or heterodimers, combines with phosphorylated proteins that regulate various metabolic processes to form a complex, regulates the activity of phosphorylated proteins, and plays an important role in plant stress and signal transduction^[50]. In this study, 33 GF14 protein family members were identified in allotetraploid *G. hirsutum*((AD)₁), which is consistent with existing reports in plants, such as rice(8)^[4], *Arabidopsis*(13)^[5], tomato(12)^[51], grape(11)^[8], poplar(12)^[9] and other species, are the species with the most GF14 family members.

G. arboreum(A₂), *G. raimondii*(D₅) and *G. barbadense*((AD)₂) have the closest relatives to *G. hirsutum*((AD)₁). It is of great significance to study the evolutionary transmission between cotton species^[20,21,24] for understanding the extension of GhGF14 protein family in cotton. Different from other species, *G. hirsutum*((AD)₁), an allotetraploid, has undergone polyploidization, which exists two subgroups of AD genes. This study shows that in *G. hirsutum*((AD)₁), there are 34 pairs of paralogous genes with collinearity between AD groups. This proves that the evolutionary amplification of GF14 protein is relatively conserved among cotton AD gene subgroups, and is also related to the polyploidization experienced by *G. hirsutum*((AD)₁); Among the four major cotton species, there is only one dense tandem repeat segment on the AD

subgroup chromosomes of *G. arboreum*(A₂), *G. raimondii*(D₅), *G. barbadense*((AD)₂) and *G. hirsutum*((AD)₁), and the remaining repeat segments are roughly the same. This indicates that the evolution and expansion of GF14 protein is slower, which shows a certain higher homology between cotton AD gene subgroups.

About 200 million years ago, when angiosperms split into monocotyledonous and dicotyledonous plants, there were two subgroups in the GF14 protein family, the ϵ group and the non- ϵ group^[11]. In this study, based on the amino acid sequence, gene structure information and phylogenetic tree relationship among *G. hirsutum*((AD)₁), rice and *Arabidopsis*, the 33 GhGF14 gene members in *G. hirsutum*((AD)₁) were divided into two broadgroups: ϵ group and non- ϵ group (Fig. 4), this result is also consistent with studies in other species^[4-12]. Based on the results of collinearity analysis by the genome data of *Arabidopsis*, rice and *G. hirsutum*((AD)₁) (Fig. 3B), there are 19 pairs of orthologous genes with collinearity between *G. hirsutum*((AD)₁) and *Arabidopsis*. There is no collinearity between *G. hirsutum*((AD)₁) and rice, which on the other hand proves that the evolution of the GF14 protein family may indeed be carried out before the differentiation of angiosperms to monocots and dicots.

As a kind of functional regulatory protein, GF14 protein is mostly limited to fiber development in the previous research in cotton. For example, Shi Haiyan et al.^[52] found that the Gh14-3-3L gene is mainly expressed in the early stage of fiber development, and reaches a peak values 10 days after anthers stage, suggesting that this gene may be involved in regulating fiber elongation; Zhang Zeting et al.^[53] found that six Gh14-3-3 proteins in cotton may be preferentially expressed in fibers and participate in the regulation of fiber cell elongation; Zhou Ying et al.^[54] found that cotton 14-3-3 protein can participate in the regulation of fiber initiation and elongation by regulating brassinolide signal transduction. In this study, we specifically found that in the GhGF14 gene family, there are the same 10 gene members (GhGF14-09, 25, 05, 21, 01, 14, 23, 04, 20, 30), Specific expression under cold, heat, drought and salt stress. This indicates that GF14 protein may also play an important role in cotton's response to abiotic stress, therefore, these genes may be the main functional genes in the GF14 family. It is also surprising that most of these gene members belong to the non- ϵ subgroup of the GhGF14 family, which indicates that there may be some structure in the non- ϵ subgroup gene sequence to make it function.

The research on the function of GF14 protein in plants is more in model plants. In these studies, it was found that GF14 protein can interact with a series of drought stress response-related proteins, enzymes or hormones, such as ion channel protein^[55], plasma membrane H⁺-ATPase^[56], ABA^[57], etc. The results of the qRT-PCR experiment in this study showed (Fig. 8) that the seven genes from GhGF14 gene family under drought stress in cotton, as the stress time increases, made significant responses in different upland cotton materials. This indicates that the GhGF14 protein family may interact with certain transcription factors or kinases, and have a negative regulatory effect on drought stress; The tissue-specific results showed (Fig. 9) that with the increase of stress

time, the relative expression levels of the seven GhGF14 genes in the roots and stems showed a downward-regulated trend, but all showed an upward-regulated trend in the leaves. This also indicates that GhGF14 protein may be regulated by a certain mechanism under drought stress, and decrease or increase in roots or leaves to achieve the purpose of inhibiting or enhancing the transport and synthesis of certain drought-responsive proteins or hormones. So as to complete the mechanism of response to drought stress. This also indirectly indicates that the GhGF14 protein family may indeed participate in the drought stress response mechanism of plants, but further research is needed.

Conclusion

Through phylogenetic tree and homology analysis, explored that the evolution and expansion of the GF14s protein family in the two major gene subgroups of cotton AD are highly conserved. And according to the evolutionary relationship between rice, *Arabidopsis* and *G. hirsutum*((AD)₁), the 33 GhGF14 protein family members are systematically divided into two GF14 subgroups, the epsilon group and the non- epsilon group. RNA-seq expression pattern analysis, it is indicated that this protein family may be involved in regulating multiple anti-stress response mechanisms in cotton. The prediction of the promoter region and qRT-PCR analysis, explored that it may have important regulatory functions in the process of drought stress response. Therefore, these results provides an important scientific basis for further research on the anti-stress function of the GF14 protein family in cotton, especially the regulation mechanism in response to drought stress, and also provides a potential help for further improving the adaptability and yield of cotton.

Methods

Plant material

In this study, Xinjiang popularized *Gossypium hirsutum*((AD)₁) resource materials KK1543 (drought-resistance materials, Xinjiang Academy of Agricultural Sciences) and Xinluzao 26 (drought-sensitive materials, Xinjiang Bazhou Agricultural Science Research Institute)^[22] were used as experimental materials. 1/2 Hoagland nutrient solution hydroponics was applied, 12 h light/12 h dark, 25 °C for greenhouse culture. When cotton seedlings are cultivated to the three-leaf stage, they are treated with 15% PEG₆₀₀₀ solution^[23] to simulate drought stress. Three cotton seedlings with the same growing vigor were selected at 0h, 1h, 3h, 6h, 12h, 24h, 48h, 72h, and the roots, stems, leaves and other tissues were sampled and stored in liquid nitrogen.

Identification of GF14 protein family

Using the newly published genome sequence as a reference background^[20,21,24-26], all sequence information processing is done by Excel worksheet and TBtools^[27]. The reference genome and proteome files of *Gossypium arboreum*(A₂), *G. raimondii*(D₅),

G. hirsutum((AD)₁) and *G. barbadense*((AD)₂) are all downloaded from the CottonFGD database^[28](<https://cottonfgd.org/>). The reference genome and proteome files of *Arabidopsis* and rice are all downloaded from EnsemblPlants database^[29](<http://plants.ensembl.org/>). Firstly, with the help of the Hidden Markov Model(HMM)(pfam00244) of the conserved domains of GF14 protein, the Blast alignment was performed in the local proteomic database to obtain the protein sequence with the specific domains of GF14 protein; Secondly, verify the GF14 family-specific domains in the NCBI-CDD database^[30](<https://www.ncbi.nlm.nih.gov/cdd>) for all sequences that are presumed to be GF14 protein from the comparison, and remove the sequences that do not contain the GF14 protein domain for further analysis; Finally, use the ProtParam online tool(<https://web.expasy.org/protparam/>) in the ExPASy^[31] online website to analyze the molecular weight (MW), isoelectric point (PI) and other physical and chemical properties of the GhGF14 protein sequence, and use the ProtComp9.0 online tool on Softberry(<http://www.softberry.com>) to predict the subcellular localization of these gene members.

Evolutionary tree analysis

The GF14 family protein sequences identified from cotton, rice^[4], and *Arabidopsis*^[5] were used to perform ClustalW multiple sequence alignment, and the evolution was constructed using the Neighbor-Joining(NJ) method in MEGA-X^[32] Tree, check that the parameter bootstrap value is set to 1000. Finally, use the online website Evolview^[33](<https://evolgenius.info/evolview-v2/>) and Adobe Illustrator CS6 to beautify it.

Homology identification and collinearity analysis

According to CottonFGD^[28], the cotton genome and chromosome database, TBtools^[27] showed the homology of GF14 protein family members on cotton chromosomes. Through the CottonGen^[34] database(<https://www.cottongen.org/>), Blast^[35], TBtools^[27] and MCScanX^[36], the homology of the GF14 protein family among the AD gene subgroups in *Gossypium hirsutum*((AD)₁) and between *G. hirsutum*((AD)₁) and rice and *Arabidopsis*, Carried out homology comparison and collinearity analysis, and visualized by TBtools^[27] and then used Adobe Illustrator CS6 for beautification adjustment.

Promoter region and gene structure analysis

The upstream 2000bp sequence of GhGF14 family genes was extracted by TBtools^[27] and used as the promoter region. Use PlantCARE^[37] online website(<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) to predict the cis-acting elements in each promoter region, and use TBtools^[27] to construct Venn diagrams based on the functional types of cis-acting elements; Predict the conservative motif of the protein sequence of GF14 family members through the MEME online website^[38](<http://meme-suite.org/tools/meme>); Finally, combined with the TM-1_V2.1.gene.gff annotation file^[20], use TBtools^[27] to locate the coding/non-coding region structure of the GhGF14 gene sequence to complete the drawing of the gene

structure map.

RNA-seq expression pattern analysis

Download the publicly available raw transcriptome data of *G. hirsutum*((AD)₁) TM-1 at seedling stage and different abiotic stress treatments (Genome sequencing project accession: [PRJNA248163](https://www.ncbi.nlm.nih.gov/sra/PRJNA248163)) from NCBI SRA (Sequence Read Archive) database(<https://www.ncbi.nlm.nih.gov/sra>). Refer to the latest genome file TM-1_genome_ZJU_v2.1, and perform standard transcriptome analysis on the original data to obtain the TPM expression level. Use TBtools^[27] software to draw the expression heatmap.

RNA extraction and qRT-PCR analysis

Using BLASTn^[35] function on the CottonGen^[34] website to compare the homologous genes of GF14 protein, and use DNAMAN (Lynnon Corporation, Canada) software to complete the design of specific fluorescent quantitative primers for related genes. According to the manufacturer's instructions, RNAprep Pure Polysaccharide Polyphenol Plant Total RNA Extraction Kit (Tiangen, Beijing, China) was used to extract total RNA from seedlings, and reverse transcription kit (abm) was used to synthesize first strand cDNA. Cotton Ubiquitin7 (UBQ7) was selected as the internal reference gene. Use Applied Biosystems™ 7500 Fast Real-Time PCR Systems to perform qRT-PCR amplification of related genes in each stress treatment period (3 technical repetitions), and use the $2^{-\Delta\Delta C_t}$ method^[39] to analyze the relative expression of related genes. Finally, choose GraphPad Prism version 8.0.1 for Windows to visualize the data.

Abbreviations

G. hirsutum: *Gossypium hirsutum*

G. barbadense: *Gossypium barbadense*

G. arboreum: *Gossypium arboreum*

G. raimondii: *Gossypium raimondii*

HMM: Hidden Markov Model

CDD: Conserved Domain Data

BLAST: Basic Local Alignment Search Tool

qRT-PCR: Quantitative real-time polymerase chain reaction

ABA: Absciscic acid

ABRE: Absciscic-acid-responsive element

NJ: Neighbor-joining

GF14: G-box factor 14

GhGF14s: *Gossypium hirsutum*'s GF14 gene

Ethics declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

RNA-seq of *G. hirsutum*((AD)₁) TM-1 at seedling stage and different abiotic stress treatments are available in the NCBI Sequence Read Archive (SRA) database under the accession number PRJNA248163. All other datasets supporting the results of this article are included within the article and its Additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

Major Special Project of Xinjiang Uygur Autonomous Region (2020A01002-2)

Authors' contributions

C.Q.J. contributed to the conception of the study; G.W.J. and C.Q. have equal contributions to this study; G.W.J., C.Q., Q.Y.Y., and C.Q.J. drafted and revised the manuscript. G.W.J., C.Q., Z.J.Y., W.P., Z.K., and W.Y.X. contributed to the data analysis. L.D.L., L.Y.L., J.Y., G.S.W., and S.X.N. contributed to the experiment. All authors reviewed and approved the final manuscript.

Acknowledgments

Not applicable.

Author information

Wen-Ju Gao and Qin Chen contributed equally to this work.

Affiliations

Engineering Research Centre of Cotton, Ministry of Education/College of

Agriculture, Xinjiang Agricultural University, 311 Nongda East Road, Urumqi, 830052, China

Wen-Ju Gao, Qin Chen, Jie-Yin Zhao, Peng Wang, Duo-Lu Li, Kai Zheng, Yi-Lei Long, Yang Jiao, Yu-Xiang Wang, Shi-Wei Geng, Xue-Ning Su, Yan-Ying Qu & Quan-Jia Chen

Post-doctoral Research Station of Grass Science, Xinjiang Agricultural University, 311 Nongda East Road, Urumqi, 830052, China

Qin Chen

Corresponding author

Correspondence to [Quan-Jia Chen](#).

Supplementary information

Additional files 1: Figure S1. NCBI-CDD Domain identification

Additional files 2: Table S1. Protein sequence of GF14 in different species

Additional file 3: Table S2. qRT-PCR primers

References

- [1] Moore B W. Specific acidic proteins of the nervous system[J]. *Physiological and Biochemical Aspects of Nervous Integration.*, 1967: 343-359.
- [2] Lu G, Delisle A J, De Vetten N C, et al. Brain proteins in plants: an Arabidopsis homolog to neurotransmitter pathway activators is part of a DNA binding complex[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 1992, 89(23): 11490-11494.
- [3] Lu G, Rooney M F, Wu K, et al. Five cDNAs encoding Arabidopsis GF14 proteins[J]. *Plant physiology*, 1994, 105(4): 1459-1460.
- [4] Yao Y, Du Y, Jiang L, et al. Molecular analysis and expression patterns of the 14-3-3 gene family from *Oryza sativa*[J]. *Journal of biochemistry and molecular biology*, 2007, 40(3): 349-357.
- [5] Rosenquist M, Alsterfjord M, Larsson C, et al. Data mining the Arabidopsis genome reveals fifteen 14-3-3 genes. Expression is demonstrated for two out of five novel genes[J]. *Plant physiology*, 2001, 127(1): 142-149.
- [6] Qin C, Cheng L, Shen J, et al. Genome-Wide Identification and Expression Analysis of the 14-3-3 Family Genes in *Medicago truncatula*[J]. *Frontiers in plant science*, 2016, 7: 320.
- [7] Wang Y, Ling L, Jiang Z, et al. Genome-wide identification and expression analysis of the gene family in soybean (*Glycine max*)[J]. *PeerJ*, 2019, 7: e7950.
- [8] Cheng C, Wang Y, Chai F, et al. Genome-wide identification and characterization of the 14-3-3 family in *Vitis vinifera* L. during berry development and cold- and heat-stress response[J]. *BMC genomics*, 2018, 19(1): 579.
- [9] Tian F, Wang T, Xie Y, et al. Genome-wide identification, classification, and expression analysis of 14-3-3 gene family in *Populus*[J]. *PloS one*, 2015, 10(4): e0123225.
- [10] Wu K, Rooney M F, Ferl R J. The Arabidopsis 14-3-3 multigene family[J]. *Plant physiology*, 1997, 114(4): 1421-1431.
- [11] Wang W, Shakes D C. Molecular evolution of the 14-3-3 protein family[J]. *Journal of molecular evolution*, 1996, 43(4): 384-398.
- [12] Rosenquist M, Sehnke P, Ferl R J, et al. Evolution of the 14-3-3 protein family: does the large number of isoforms in multicellular organisms reflect functional specificity?[J]. *Journal of molecular evolution*, 2000, 51(5): 446-458.
- [13] Jahn T, Fuglsang A T, Olsson A, et al. The 14-3-3 protein interacts directly with the C-terminal region of the plant plasma membrane H(+)-ATPase[J]. *The Plant cell*, 1997, 9(10): 1805-1814.
- [14] Igarashi D, Ishida S, Fukazawa J, et al. 14-3-3 proteins regulate intracellular localization of the bZIP transcriptional activator RSG[J]. *Plant Cell*, 2001, 13(11): 2483-97.
- [15] Ishida S, Fukazawa J, Yuasa T, et al. Involvement of 14-3-3 signaling protein binding in the functional regulation of the transcriptional activator REPRESSION OF SHOOT GROWTH by gibberellins[J]. *The Plant cell*, 2004, 16(10): 2641-2651.
- [16] Ren Y-R, Yang Y-Y, Zhang R, et al. MdGRF11, an apple 14-3-3 protein, acts as a positive regulator of drought and salt tolerance[J]. *Plant science : an international journal of experimental plant biology*, 2019, 288: 110219.
- [17] Zhang Y, Zhao H, Zhou S, et al. Expression of TaGF14b, a 14-3-3 adaptor protein gene from wheat, enhances drought and salt tolerance in transgenic tobacco[J]. *Planta*, 2018, 248(1): 117-137.
- [18] Visconti S, D'ambrosio C, Fiorillo A, et al. Overexpression of 14-3-3 proteins enhances cold tolerance and increases levels of stress-responsive proteins of Arabidopsis plants[J]. *Plant science : an*

international journal of experimental plant biology, 2019, 289: 110215.

[19] Yan J, He C, Wang J, et al. Overexpression of the Arabidopsis 14-3-3 protein GF14 lambda in cotton leads to a "stay-green" phenotype and improves stress tolerance under moderate drought conditions[J]. Plant & cell physiology, 2004, 45(8): 1007-1014.

[20] Hu Y, Chen J, Fang L, et al. Gossypium barbadense and Gossypium hirsutum genomes provide insights into the origin and evolution of allotetraploid cotton[J]. Nat Genet, 2019, 51(4): 739-748.

[21] Huang G, Wu Z, Percy R G, et al. Genome sequence of Gossypium herbaceum and genome updates of Gossypium arboreum and Gossypium hirsutum provide insights into cotton A-genome evolution[J]. Nature genetics, 2020, 52(5): 516-524.

[22] Fenglei S, YanYing Q, QuanJia C, et al. Comprehensive evaluation of cotton drought tolerance indexes and gray relational analysis[J]. Agricultural Research in the Arid Areas, 2019, 37(01): 233-239.

[23] YaMeng L. Preliminary Evaluation on Drought Tolerance of Wild Upland Cotton [D]. Hebei Agricultural University, 2019.

[24] Wang K, Wang Z, Li F, et al. The draft genome of a diploid cotton Gossypium raimondii[J]. Nature genetics, 2012, 44(10): 1098-1103.

[25] Cheng C-Y, Krishnakumar V, Chan A P, et al. Araport11: a complete reannotation of the Arabidopsis thaliana reference genome[J]. The Plant Journal, 2017, 89(4): 789-804.

[26] Kawahara Y, De La Bastide M, Hamilton J P, et al. Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data[J]. Rice (New York, N.Y.), 2013, 6(1): 4.

[27] Chen C, Chen H, Zhang Y, et al. TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data[J]. Mol Plant, 2020, 13(8): 1194-1202.

[28] Zhu T, Liang C, Meng Z, et al. CottonFGD: an integrated functional genomics database for cotton[J]. BMC Plant Biol, 2017, 17(1): 101.

[29] Howe K L, Contreras-Moreira B, De Silva N, et al. Ensembl Genomes 2020-enabling non-vertebrate genomic research[J]. Nucleic Acids Res, 2020, 48(D1): D689-D695.

[30] Lu S, Wang J, Chitsaz F, et al. CDD/SPARCLE: the conserved domain database in 2020[J]. Nucleic Acids Res, 2020, 48(D1): D265-D268.

[31] Gasteiger E, Hoogland C, Gattiker A, et al.: Protein Identification and Analysis Tools on the ExPASy Server, Walker J M, editor, The Proteomics Protocols Handbook, Totowa, NJ: Humana Press, 2005: 571-607.

[32] Kumar S, Stecher G, Li M, et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms[J]. Mol Biol Evol, 2018, 35(6): 1547-1549.

[33] Subramanian B, Gao S, Lercher M J, et al. Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees[J]. Nucleic Acids Res, 2019, 47(W1): W270-W275.

[34] Yu J, Jung S, Cheng C-H, et al. CottonGen: a genomics, genetics and breeding database for cotton research[J]. Nucleic acids research, 2014, 42(Database issue): D1229-D1236.

[35] Altschul S F, Gish W, Miller W, et al. Basic local alignment search tool[J]. Journal of molecular biology, 1990, 215(3): 403-410.

[36] Wang Y, Tang H, Debarry J D, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity[J]. Nucleic acids research, 2012, 40(7): e49.

[37] Lescot M, Déhais P, Thijs G, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences[J]. Nucleic acids research, 2002, 30(1): 325-327.

- [38] Bailey T L, Boden M, Buske F A, et al. MEME SUITE: tools for motif discovery and searching[J]. *Nucleic Acids Res*, 2009, 37(Web Server issue): W202-8.
- [39] Livak K J, Schmittgen T D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method[J]. *Methods (San Diego, Calif.)*, 2001, 25(4): 402-408.
- [40] Wang D, Pan Y, Zhao X, et al. Genome-wide temporal-spatial gene expression profiling of drought responsiveness in rice[J]. *BMC Genomics*, 2011(1).
- [41] Ngara R, Ndimba B K. Model plant systems in salinity and drought stress proteomics studies: a perspective on Arabidopsis and Sorghum[J]. *Plant biology (Stuttgart, Germany)*, 2014, 16(6): 1029-1032.
- [42] Yang C, Lu X, Ma B, et al. Ethylene signaling in rice and Arabidopsis: conserved and diverged aspects[J]. *Molecular plant*, 2015, 8(4): 495-505.
- [43] Hernandez-Garcia C M, Finer J J. Identification and validation of promoters and cis-acting regulatory elements[J]. *Plant science : an international journal of experimental plant biology*, 2014, 217-218: 109-119.
- [44] Dai X, Xu Y, Ma Q, et al. Overexpression of an R1R2R3 MYB gene, OsMYB3R-2, increases tolerance to freezing, drought, and salt stress in transgenic Arabidopsis[J]. *Plant physiology*, 2007, 143(4): 1739-1751.
- [45] Onishi M, Tachi H, Kojima T, et al. Molecular cloning and characterization of a novel salt-inducible gene encoding an acidic isoform of PR-5 protein in soybean (*Glycine max* [L.] Merr.)[J]. *Plant physiology and biochemistry : PPB*, 2006, 44(10): 574-580.
- [46] Manavella P A, Dezar C A, Ariel F D, et al. Two ABREs, two redundant root-specific and one W-box cis-acting elements are functional in the sunflower HAHB4 promoter[J]. *Plant physiology and biochemistry : PPB*, 2008, 46(10): 860-867.
- [47] Wang Z, Zhu Y, Wang L, et al. A WRKY transcription factor participates in dehydration tolerance in *Boea hygrometrica* by binding to the W-box elements of the galactinol synthase (BhGolS1) promoter[J]. *Planta*, 2009, 230(6): 1155-1166.
- [48] Germain H, Lachance D, Pelletier G, et al. The expression pattern of the *Picea glauca* Defensin 1 promoter is maintained in *Arabidopsis thaliana*, indicating the conservation of signalling pathways between angiosperms and gymnosperms[J]. *Journal of experimental botany*, 2012, 63(2): 785-795.
- [49] Yamaguchi-Shinozaki K, Shinozaki K. A novel cis-acting element in an Arabidopsis gene is involved in responsiveness to drought, low-temperature, or high-salt stress[J]. *The Plant cell*, 1994, 6(2): 251-264.
- [50] Wilson R S, Swatek K N, Thelen J J. Regulation of the Regulators: Post-Translational Modifications, Subcellular, and Spatiotemporal Distribution of Plant 14-3-3 Proteins[J]. *Frontiers in plant science*, 2016, 7: 611.
- [51] Xu W F, Shi W M. Expression profiling of the 14-3-3 gene family in response to salt stress and potassium and iron deficiencies in young tomato (*Solanum lycopersicum*) roots: analysis by real-time RT-PCR[J]. *Annals of botany*, 2006, 98(5): 965-974.
- [52] Shi H, Wang X, Li D, et al. Molecular characterization of cotton 14-3-3L gene preferentially expressed during fiber elongation[J]. *Journal of genetics and genomics = Yi chuan xue bao*, 2007, 34(2): 151-159.
- [53] Zhang Z-T, Zhou Y, Li Y, et al. Interactome analysis of the six cotton 14-3-3s that are preferentially expressed in fibres and involved in cell elongation[J]. *Journal of experimental botany*, 2010, 61(12): 3331-3344.
- [54] Zhou Y, Zhang Z-T, Li M, et al. Cotton (*Gossypium hirsutum*) 14-3-3 proteins participate in regulation of fibre initiation and elongation by modulating brassinosteroid signalling[J]. *Plant*

biotechnology journal, 2015, 13(2): 269-280.

[55] Booij, Roberts, Vogelzang, et al. 14-3-3 proteins double the number of outward-rectifying K⁺ channels available for activation in tomato cells[J]. The Plant journal : for cell and molecular biology, 1999, 20(6): 673-683.

[56] Borch J, Bych K, Roepstorff P, et al. Phosphorylation-independent interaction between 14-3-3 protein and the plant plasma membrane H⁺-ATPase[J]. Biochemical Society transactions, 2002, 30(4): 411-415.

[57] Schoonheim P J, Sinnige M P, Casaretto J A, et al. 14-3-3 adaptor proteins are intermediates in ABA signal transduction during barley seed germination[J]. The Plant journal : for cell and molecular biology, 2007, 49(2): 289-301.

Figures

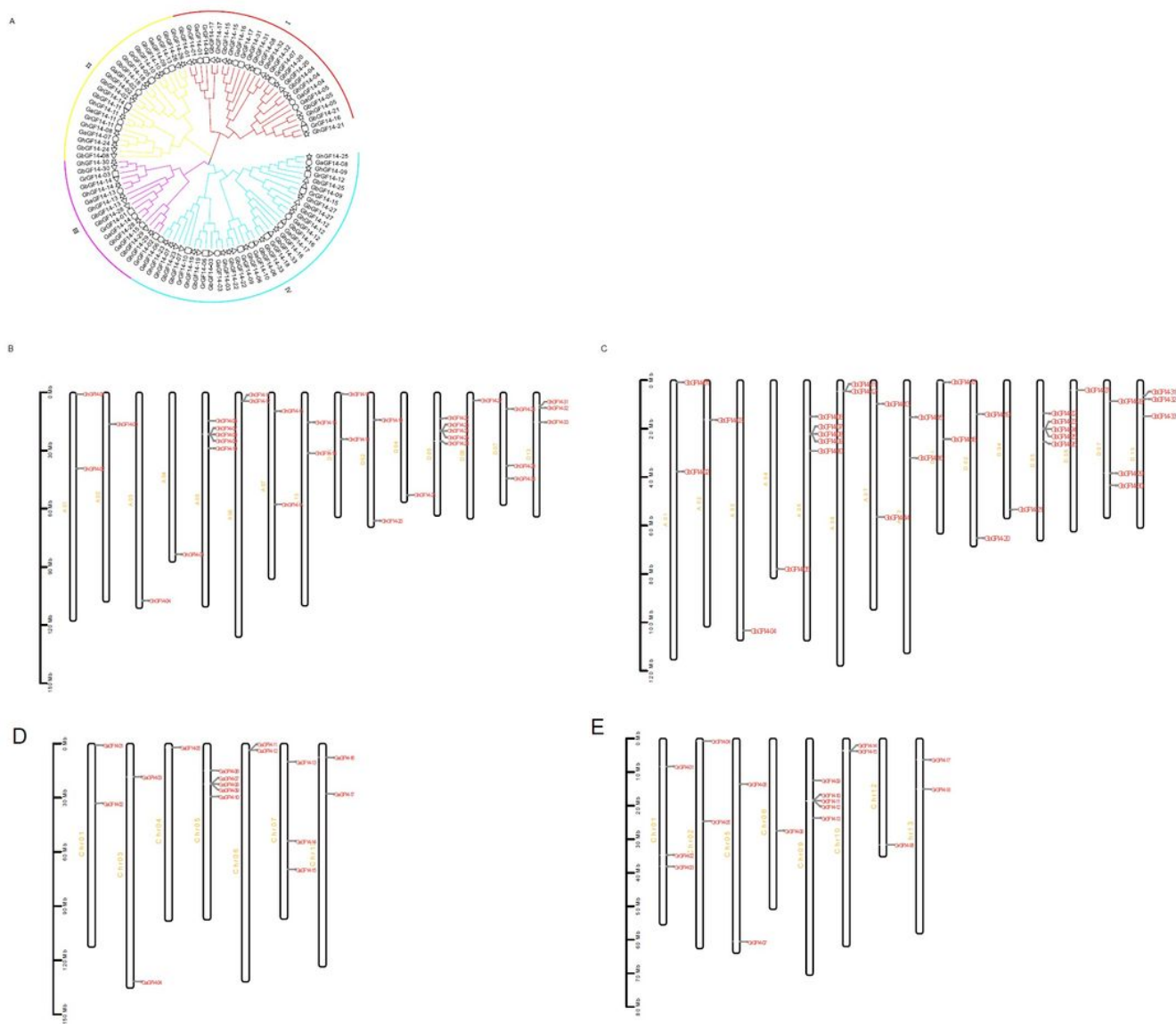


Figure 1

Chromosome location and phylogenetic tree of cotton GF14 protein family. A) NJ evolutionary tree constructed with 101 GF14 protein sequences of *Gossypium hirsutum*((AD)1), *G. barbadense*((AD)2), *G. arboreum*(A2) and *G. raimondii*(D5). The stars represent *G. hirsutum*((AD)1), triangles represent *G. barbadense*((AD)2), circles represent *G. arboreum*(A2), and squares represent *G. raimondii*(D5). (B) Chromosome location of GF14 gene in *G. hirsutum*((AD)1). (C) Chromosome location of GF14 gene in *G. barbadense*((AD)2). (D) Chromosome location of GF14 gene in *G. arboreum*(A2). (E) Chromosome location of GF14 gene in *G. raimondii*(D5).

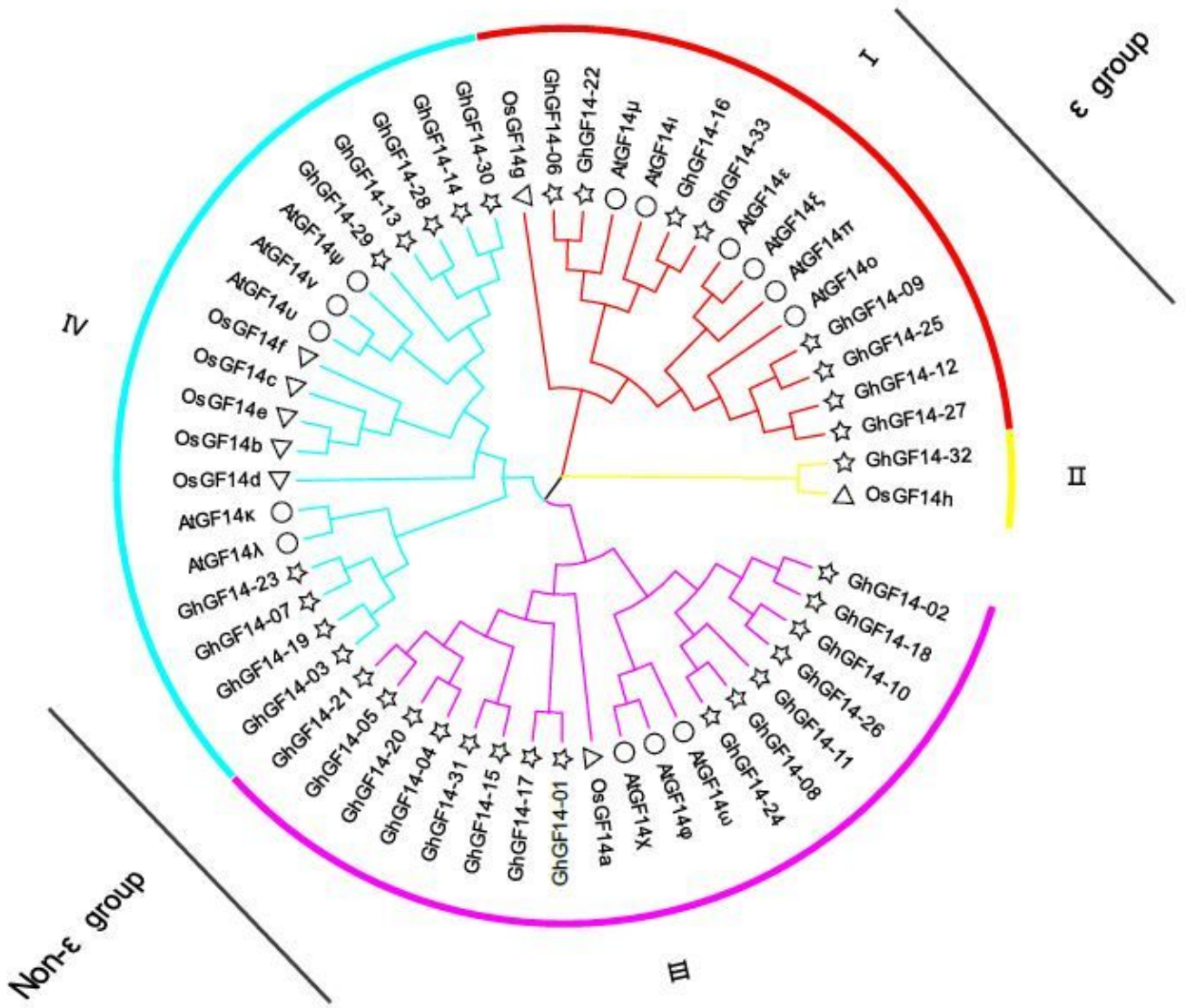
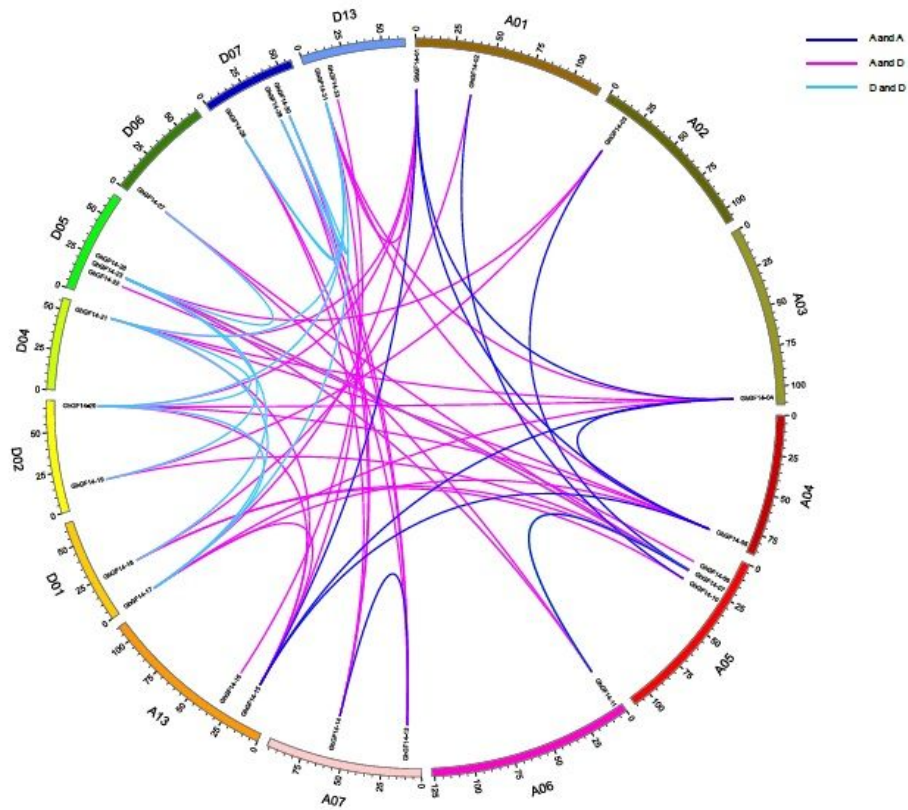


Figure 2

Phylogenetic tree of GF14 protein gene family in *Gossypium hirsutum*((AD)1), *Arabidopsis* and rice. The stars represent GF14s in *G. hirsutum*((AD)1). The circle represents GF14s in *Arabidopsis*. The triangles represent GF14s in rice. The names of GF14s in *Arabidopsis* and rice are derived from previous studies.

A



B

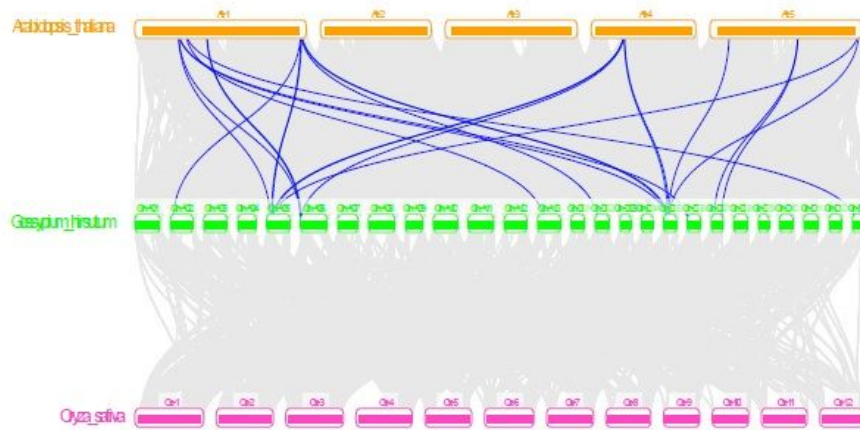


Figure 3

Homology identification and analysis of GhGF14s gene. (A) The homology of GhGF14s between the chromosomes in the AD group of *G. hirsutum*((AD)1). (B) The homology of GhGF14s among the genomes of *G. hirsutum*((AD)1), rice and Arabidopsis.

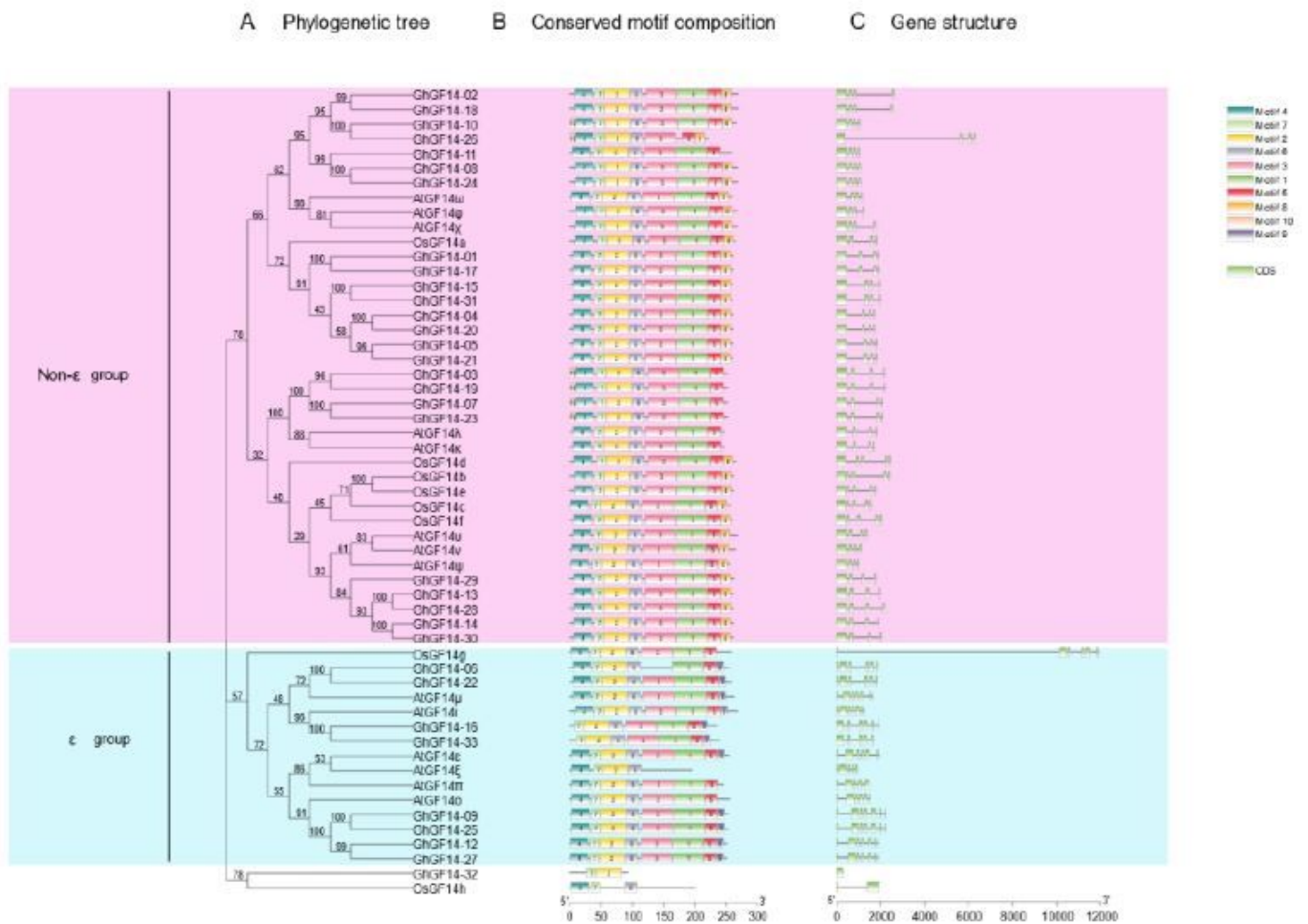


Figure 4

Gene structure diagram of GF14 protein family in *G. hirsutum*((AD)1), rice and Arabidopsis. (A) Phylogenetic tree of GF14s protein family in *G. hirsutum*((AD)1), rice and Arabidopsis. (B) Conservative motif structure prediction. (C) The distribution of exons and introns in the CDS region of GF14s.

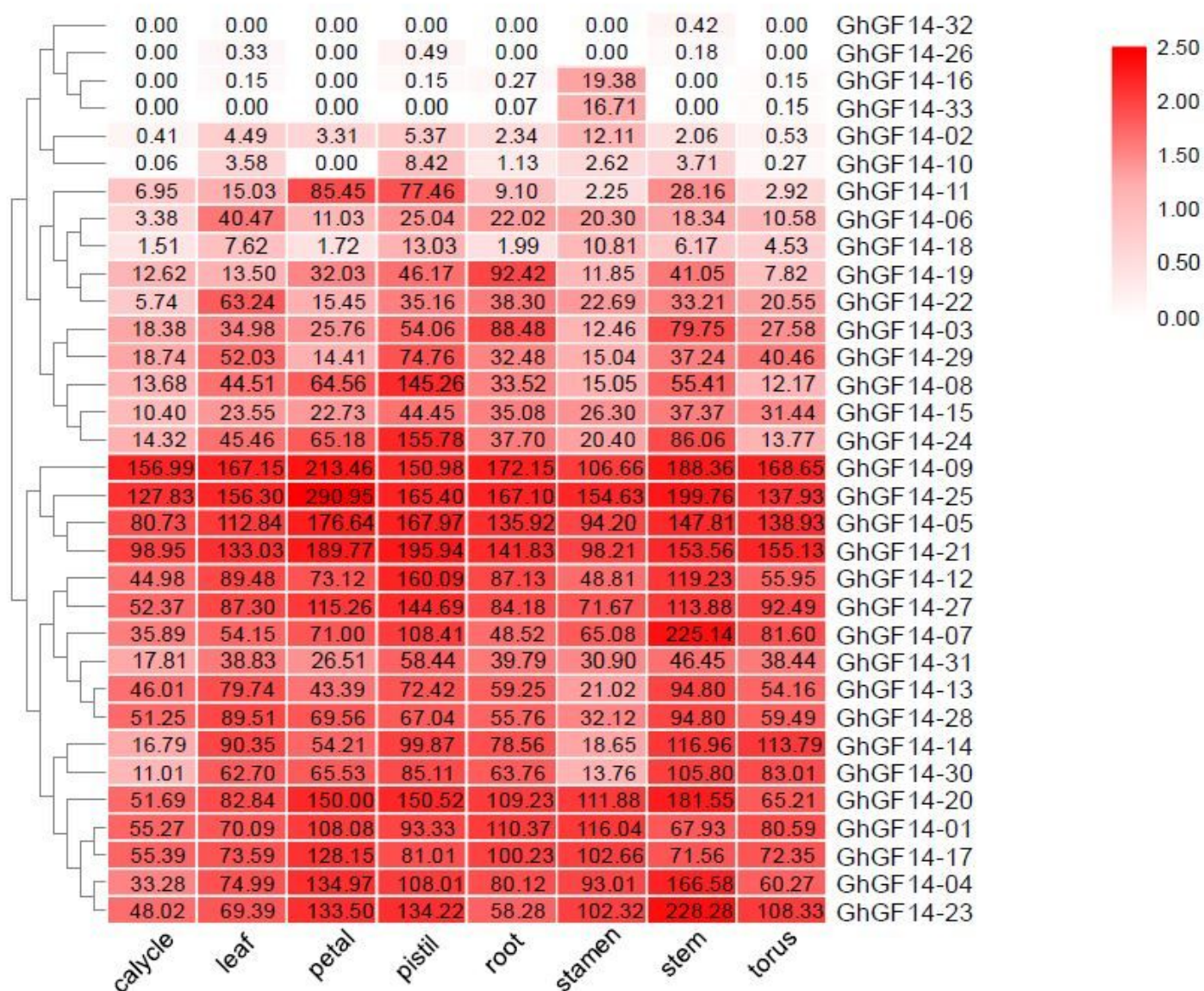


Figure 5

Tissue specificity at the transcriptome level of the GhGF14 gene family. Use $\log_{10}(x+1)$ standardization for data processing, x is the original expression value, and the original expression value is shown in the figure; the red to white gradient represents the change of expression from high to low.

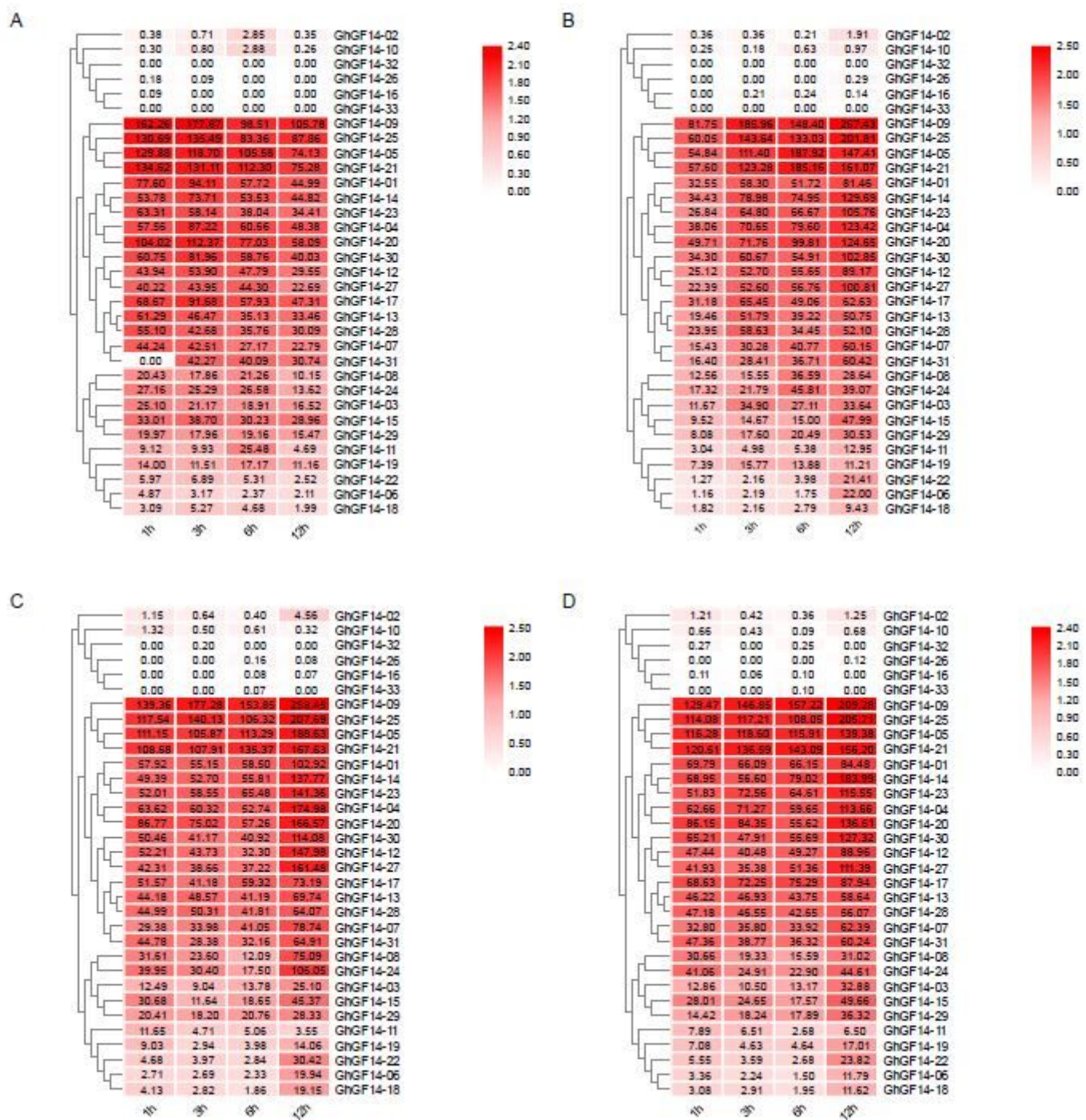
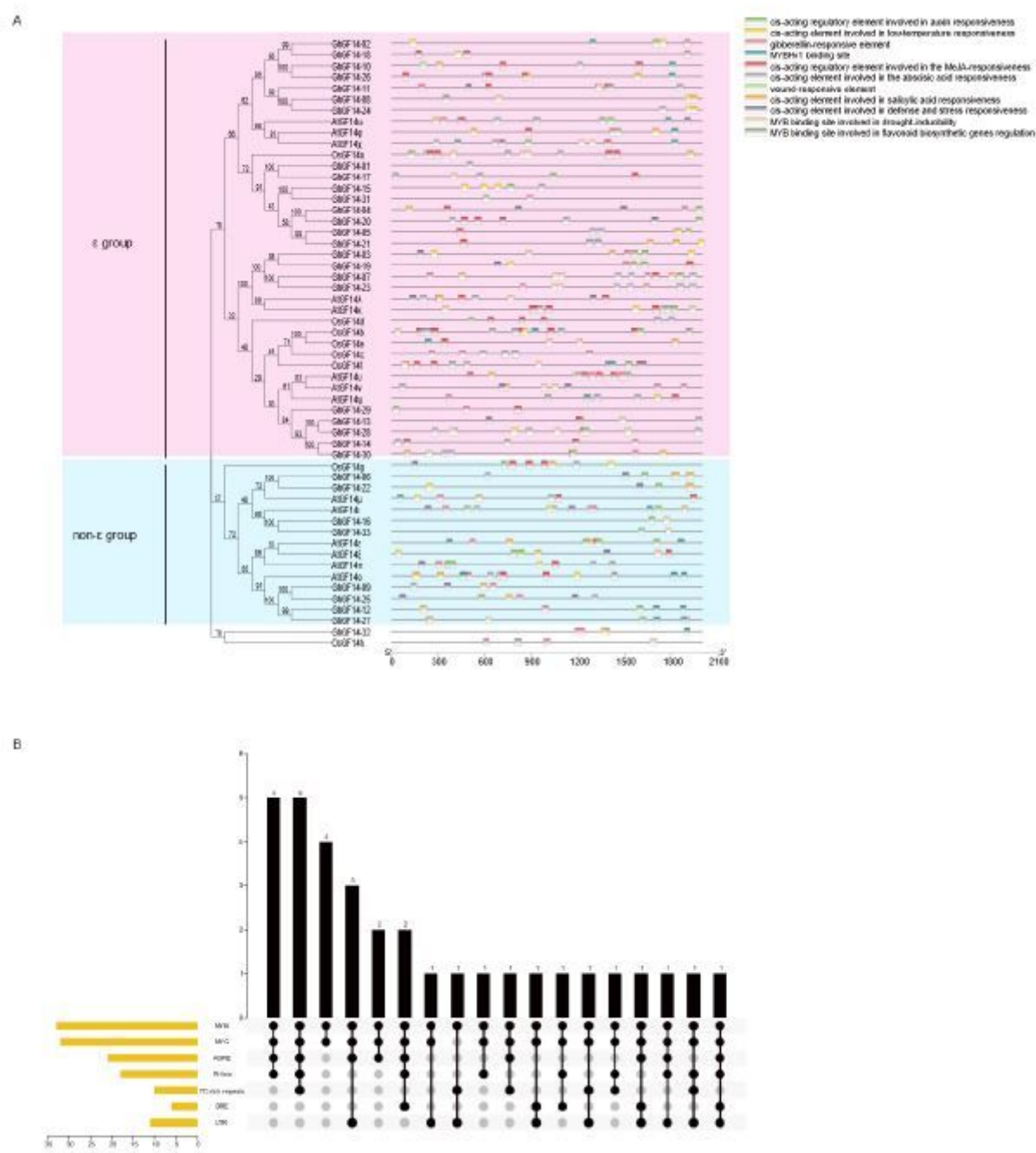


Figure 6

Analysis of the expression pattern of GhGF14s gene family. (A) Changes in GhGF14s gene expression during 1h, 3h, 6h, 12h periods under cold stress; (B) Changes in GhGF14s gene expression during 1h, 3h, 6h, 12h periods under heat stress; (C) Changes in GhGF14s gene expression during 1h, 3h, 6h and 12h periods under drought stress; (D) Changes in GhGF14s gene expression during 1h, 3h, 6h and 12h periods under salt stress. Use $\log_{10}(x+1)$ standardization for data processing, x is the original expression value,

and the original expression value is shown in the figure; the red to white gradient represents the change of expression from high to low.



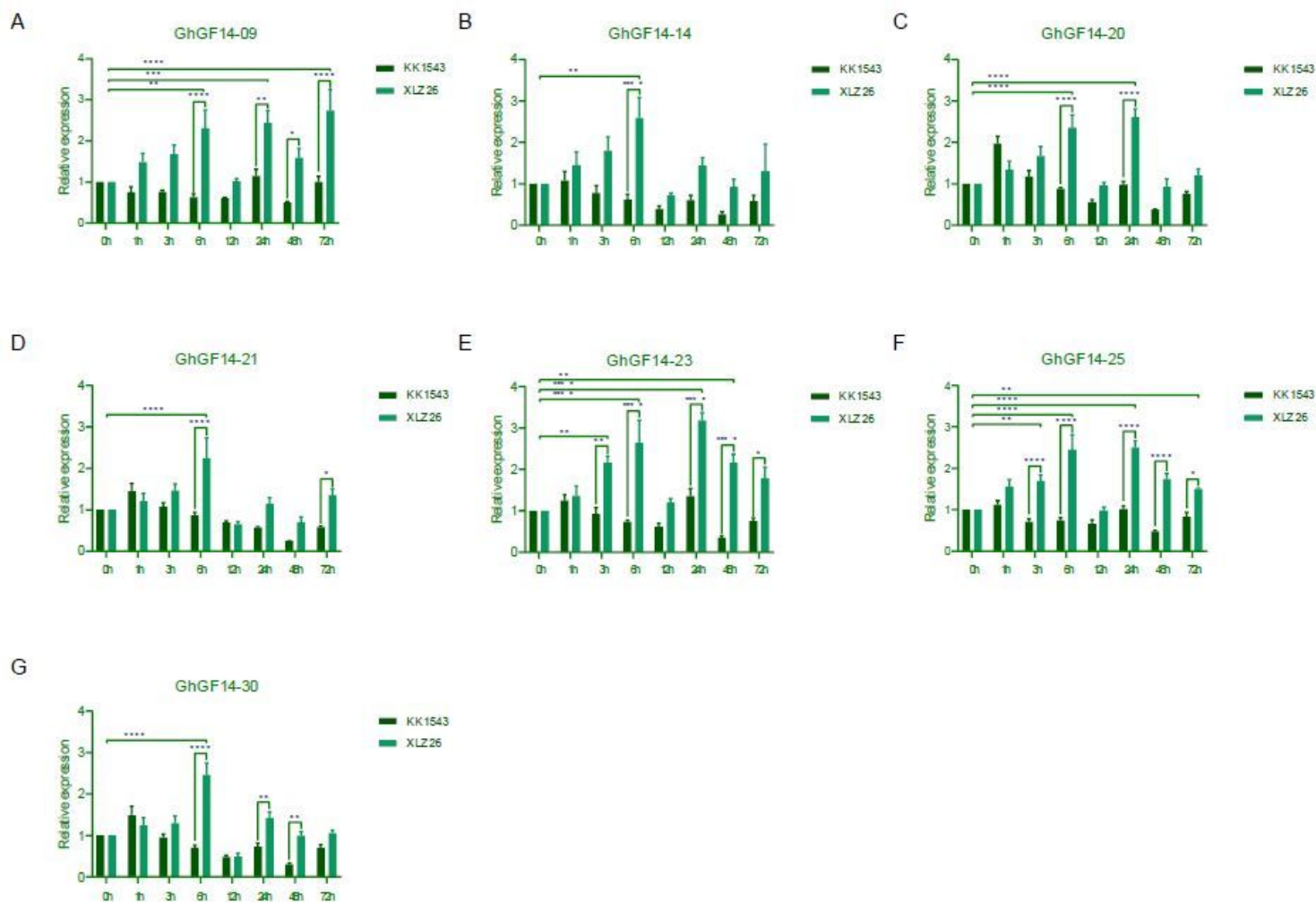


Figure 8

Drought stress qTR-PCR analysis. ABCDEFG respectively represents the corresponding gene name. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

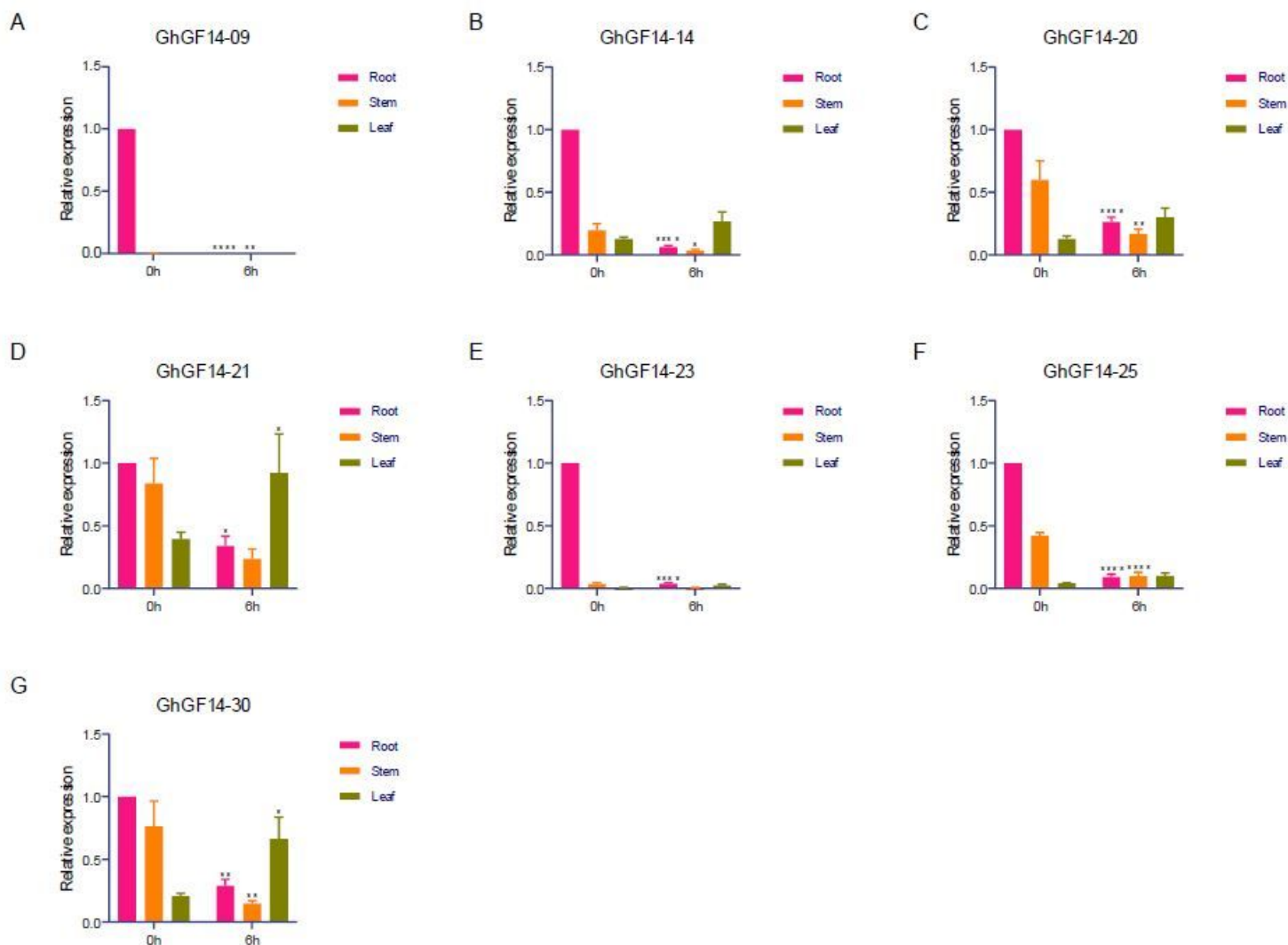


Figure 9

QTR-PCR tissue specificity analysis. ABCDEFG respectively represents the corresponding gene name. * P<0.05, ** P<0.01, *** P<0.001, **** P<0.0001.

Supplementary Files

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

- [FigureS1.NCBICDDDomainidentification.pdf](#)
- [TableS1.ProteinsequenceofGF14indifferentspecies.xlsx](#)
- [TableS2.qRTPCRprimers.xlsx](#)