

RESEARCH ARTICLE

Open Access



Genome-wide identification of the DUF668 gene family in cotton and expression profiling analysis of GhDUF668 in *Gossypium hirsutum* under adverse stress

Jieyin Zhao, Peng Wang, Wenju Gao, Yilei Long, Yuxiang Wang, Shiwei Geng, Xuening Su, Yang Jiao, Quanjia Chen and Yanying Qu*

Abstract

Background: Domain of unknown function 668 (DUF668) may play a crucial role in the plant growth and developmental response to adverse stress. However, our knowledge of the function of the DUF668 gene family is limited.

Results: Our study was conducted based on the DUF668 gene family identified from cotton genome sequencing. Phylogenetic analysis showed that the DUF668 family genes can be classified into four subgroups in cotton. We identified 32 DUF668 genes, which are distributed on 17 chromosomes and most of them located in the nucleus of *Gossypium hirsutum*. Gene structure and motif analyses revealed that the members of the DUF668 gene family can be clustered in *G. hirsutum* into two broad groups, which are relatively evolutionarily conserved. Transcriptome data analysis showed that the GhDUF668 genes are differentially expressed in different tissues under various stresses (cold, heat, drought, salt, and *Verticillium dahliae*), and expression is generally increased in roots and stems. Promoter and expression analyses indicated that *Gh_DUF668-05*, *Gh_DUF668-08*, *Gh_DUF668-11*, *Gh_DUF668-23* and *Gh_DUF668-28* in *G. hirsutum* might have evolved resistance to adverse stress. Additionally, qRT-PCR revealed that these 5 genes in four cotton lines, KK1543 (drought resistant), Xinluzao 26 (drought sensitive), Zhongzhimian 2 (disease resistant) and Simian 3 (susceptible), under drought and *Verticillium* wilt stress were all significantly induced. Roots had the highest expression of these 5 genes before and after the treatment. Among them, the expression levels of *Gh_DUF668-08* and *Gh_DUF668-23* increased sharply at 6 h and reached a maximum at 12 h under biotic and abiotic stress, which showed that they might be involved in the process of adverse stress resistance in cotton.

Conclusion: The significant changes in *GhDUF668* expression in the roots after adverse stress indicate that *GhDUF668* is likely to increase plant resistance to stress. This study provides an important theoretical basis for further research on the function of the DUF668 gene family and the molecular mechanism of adverse stress resistance in cotton.

Keywords: Cotton, DUF668 gene family, Bioinformatics analysis, Adverse stress, Expression analysis

* Correspondence: xjyq5322@126.com

Engineering Research Centre of Cotton, Ministry of Education/College of Agriculture, Xinjiang Agricultural University, 311 Nongda East Road, Urumqi 830052, China



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Plant biologists have always been attracted to the structure, function, and evolutionary model of gene families. The interaction and adaptation between the environment and plants are well studied based on the information of these gene families [1]. Among them, the domain of unknown function (DUF) family refers to a certain protein family with unknown functions, and they play a key role in the plant response to stress [2]. In recent years, a large number of species' genomes have been sequenced, and the number of DUF superfamilies has increased rapidly. As of 2010, the entire family has expanded to DUF2607 [3]. The Pfam database (version 33.1) now includes 18,259 gene families, of which nearly 31% (5645) are composed of DUF families [4]. The rapid development of genomics and proteomics provides important bioinformatics data for the systematic study of DUF superfamily proteins and lays the foundation for the study of these DUF family genes in regulating plant growth and development and responding to biotic and abiotic stresses.

However, there have been some reports of other DUF gene families in many plants. These include the DUF221, DUF810, DUF866, DUF936 and DUF1618 gene families in rice and the DUF581 and DUF724 gene families in *Arabidopsis* [5–11]. DUF27 confers the ability to bind to ADP-ribose specifically [12]. The DUF283 domain is required for siRNA processing in gene silencing [13, 14], and the DUF538 superfamily has the ability to hydrolyze chlorophyll [15, 16]. A previous study in *Arabidopsis* showed that ESK1 (AT3g55990) of the DUF231 gene family is a new negative regulator of cold acclimation [17]. Another study showed that it inhibits the expression of ATRDUF1 and ATRDUF2 (both are RING-DUF1117 E3 ubiquitin ligases) [18]. Abscisic acid (ABA) mediates the response to drought stress. The DUF1644 gene OsSIDP366 positively regulates the response to drought and salt stress in rice [19]. Transgenic rice overexpressing OsSIDP366 shows stronger drought resistance and salt tolerance [20]. Other DUF genes have also been characterized to be related to abiotic stresses, and SIDP361 (DUF1644), OsDSR2 (DUF966) and OsDUF810 [6, 7, 19] from the DUF2275 family are regulated by nutritional status and dehydration during development. Overexpression of the salt-inducing gene TaSRHP (containing the DUF581 domain) in wild-type *Arabidopsis* can enhance its resistance to salt and drought stress [21]. The function of DUF stress tolerance is currently reported in only model plants, while comprehensive DUF gene family analysis in other plant species is rarely reported.

Although some members of the DUF gene family have been identified, a great number of DUF members are still unknown, especially in cotton. The DUF668 family

was identified as a conserved domain containing 29 amino acids. However, limited research has been conducted on this gene family. To date, the DUF668 gene family has been reported in only rice [2]. Previous studies have shown that all tetraploid cottons are directly evolved by doubling the genome after crossing the A and D genomes. Among them, the *G. arboreum* (A2-genome) is used as the donor of A genome, and the *G. hirsutum* (D5-genome) is used as the donor of D genome [22–25]. At present, all of the major cotton areas worldwide are threatened by varying degrees of salt, alkali, drought, cold damage and disease [26–30]. It has become an important scientific issue to continuously identify and screen genes with multiple stress resistance functions and develop related molecular markers in cotton research. Genome sequencing has achieved remarkable results in cotton [22–25], making it possible to systematically identify and study gene families in cotton. DUF668 family genes have shown the potential importance of participating in stress resistance in plants [2]. The evolution, function and classification of this gene family in cotton have not been systematically studied. In this study, members of the DUF668 family were systematically identified, and bioinformatic analyses were performed based on cotton genome data. Chromosome distribution, gene replication, promoter cis-acting elements, and expression profiles of the GhDUF668 gene were analyzed in different tissues and under various stresses. qRT-PCR was used to analyze the expression of candidate genes under drought and *Verticillium dahliae* (V991) treatments, revealing their possible biological functions. The results will further broaden our understanding of the roles of DUF668 genes in plants, providing a basis for further research on the functions of these genes in cotton under adverse stresses and laying a foundation for the subsequent analysis of their functions.

Results

Identification of the DUF668 gene family from cotton

To investigate the copy number variation in the DUF668 genes during cotton evolution, a comprehensive search was conducted for DUF668 genes across cotton lineages, including *G. arboreum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*. The results were verified in the NCBI-CDD database (Figure S1). In the end, there were 17, 17, 32, and 33 sequences in *G. arboreum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*, respectively (Table S1). The results showed that the numbers of DUF668 genes in *G. arboreum* and *G. raimondii* were almost similar as were those in *G. hirsutum* and *G. barbadense*. The DUF668 family genes in two diploid cotton species are basically half of the number in two tetraploid cotton species, which conforms to the known evolutionary

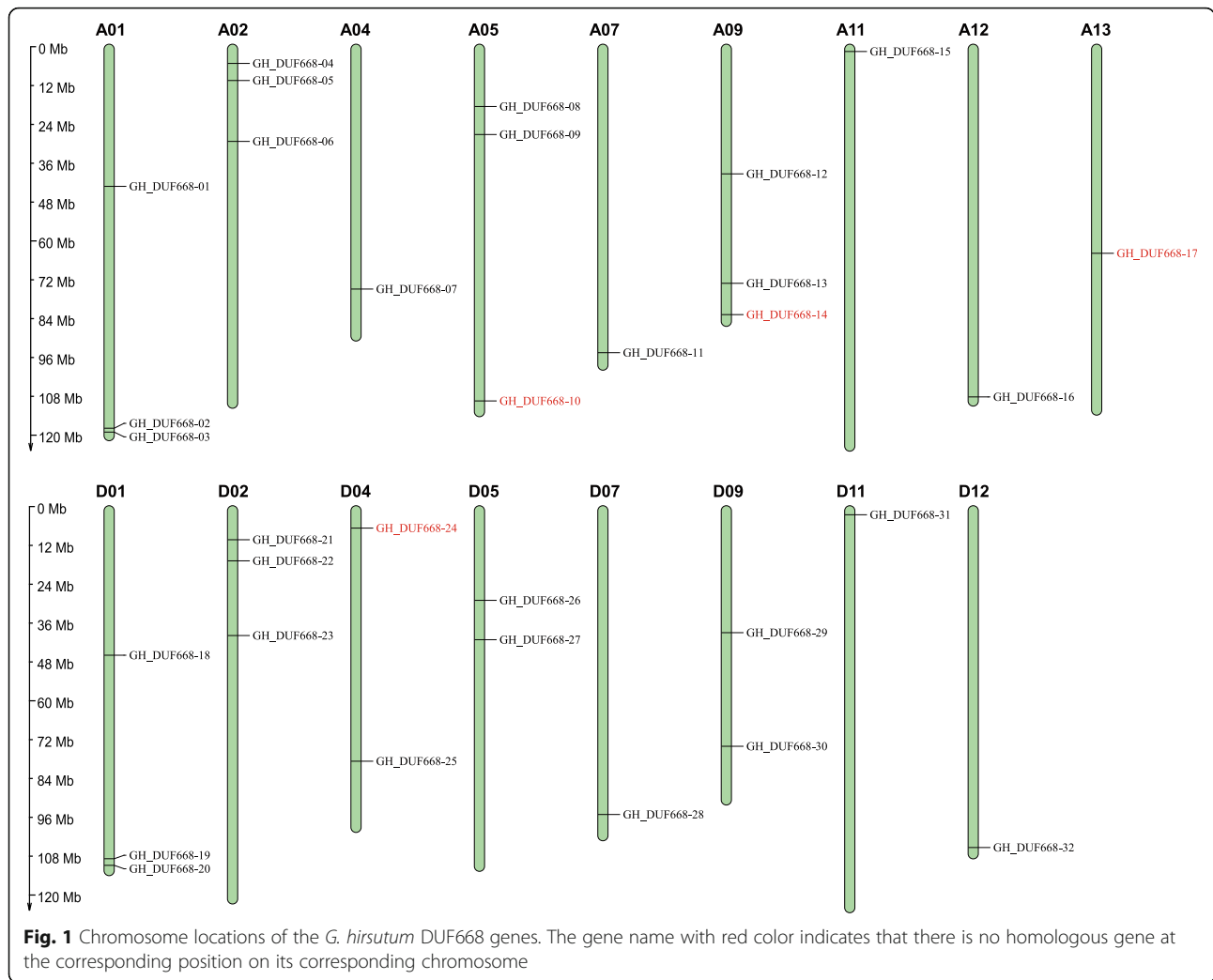
relationship of cotton [24, 25], indicating that the DUF668 family is conserved in the evolution of cotton. *Gh_DUF668-01* ~ *Gh_DUF668-32* were named according to the position of the 32 sequences on the chromosome (Table 1) in *G. hirsutum*. The open reading frame (ORF) of the DUF668 family genes in *G. hirsutum* is 630 ~ 1959 bp in length, and the encoded protein contains 209 ~ 652 amino acid residues. The relative molecular mass is between 23.46 and 72.69 kDa, and the theoretical isoelectric point is between 5.29 and 9.83. Each of the family members

contains a DUF668 domain. The subcellular localization of proteins showed that 27 were located in the nucleus, 4 were located in the chloroplast, and 1 was located in the inner membrane.

Thirty-two GhDUF668 genes were distributed on 17 chromosomes (A01, A02, A04, A05, A07, A09, A11, A12, A13, D01, D02, D04, D05, D07, D09, D11, and D12) of *G. hirsutum* (Fig. 1). Subgroup A and subgroup D contained 17 and 15 sequences, respectively. Previous studies suggested that *G. arboreum* and *G. raimondii* were donor species for subgenome A and subgenome D,

Table 1 Information on the DUF668 gene family in *G. hirsutum*

Gene name	Gene ID	Open reading frame/bp	Protein length/aa	Relative molecular weight (r)/kDa	Theoretical isoelectric point (pI)	Subcellular localization
GhDUF668-01	GH_A01G1304	1875	598	67.11	9.28	nucleus
GhDUF668-02	GH_A01G2268	1707	568	64.86	9.22	nucleus
GhDUF668-03	GH_A01G2392	1392	463	52.72	9.2	chloroplast
GhDUF668-04	GH_A02G0391	1788	595	67.2	6.95	nucleus
GhDUF668-05	GH_A02G0661	1959	652	72.62	9.45	nucleus
GhDUF668-06	GH_A02G0964	1113	370	41.94	9.72	nucleus
GhDUF668-07	GH_A04G1034	1539	512	58.73	9.36	nucleus
GhDUF668-08	GH_A05G1859	1854	617	68.53	9.49	nucleus
GhDUF668-09	GH_A05G2503	1947	648	71.88	9.29	nucleus
GhDUF668-10	GH_A05G4150	1758	585	65.7	9.16	nucleus
GhDUF668-11	GH_A07G2347	1875	624	69.82	8.99	nucleus
GhDUF668-12	GH_A09G0607	1353	450	51.07	9.23	nucleus
GhDUF668-13	GH_A09G1530	1617	538	61.45	9.83	nucleus
GhDUF668-14	GH_A09G2457	1782	593	66.63	8.31	nucleus
GhDUF668-15	GH_A11G0076	1386	461	51.95	9.43	chloroplast
GhDUF668-16	GH_A12G2838	1179	392	44.97	9.24	chloroplast
GhDUF668-17	GH_A13G1138	1455	484	54.28	8.96	nucleus
GhDUF668-18	GH_D01G1376	1797	598	67.05	9.2	nucleus
GhDUF668-19	GH_D01G2352	1707	568	64.73	9.27	nucleus
GhDUF668-20	GH_D01G2470	1392	463	52.66	9.01	chloroplast
GhDUF668-21	GH_D02G0412	1791	596	67.2	7.55	nucleus
GhDUF668-22	GH_D02G0670	1959	652	72.69	9.43	nucleus
GhDUF668-23	GH_D02G1007	1836	611	68.02	9.59	nucleus
GhDUF668-24	GH_D04G0230	630	209	23.46	5.29	nucleus
GhDUF668-25	GH_D04G1367	1551	516	59.29	9.36	nucleus
GhDUF668-26	GH_D05G1897	1854	617	68.75	9.56	nucleus
GhDUF668-27	GH_D05G2525	1947	648	71.93	9.16	nucleus
GhDUF668-28	GH_D07G2291	1875	624	69.64	9.06	nucleus
GhDUF668-29	GH_D09G0545	1356	451	51.18	9.27	nucleus
GhDUF668-30	GH_D09G1537	1617	538	61.58	9.79	nucleus
GhDUF668-31	GH_D11G0081	1386	461	52.1	9.37	chloroplast
GhDUF668-32	GH_D12G2861	852	283	32.24	9.19	endomembrane



respectively. The number of GhDUF668 genes in subgenome A was consistent with the number of GaDUF668 genes, and two of the DUF668 genes were missing from subgenome D compared to the number of GrDUF668 genes. This result indicated that subgroup D might have lost genes due to redundant gene functions during cotton evolution. Only one sequence of this family was on chromosome A04, while chromosome D04 in *G. hirsutum* contained two sequences. Three sequences were observed on chromosomes A05 and A09, while chromosomes D05 and D09 contained two sequences. The A13 chromosome contained one sequence, but the GhDUF668 gene sequence was not contained in D13 chromosome. This result showed that the DUF668 genes might have been lost and duplicated in the process of evolution. However, there was a strong correlation between subgroup A and subgroup D, which was also in line with the evolutionary relationship in cotton [22–25].

Phylogenetic analysis of the DUF668 gene family in cotton

To explore the phylogenetic relationship of the cotton DUF668 genes, a phylogenetic tree was constructed. DUF668 gene protein sequences (Table S2) from four different cotton subspecies were used. All of the DUF668 proteins can be divided into 4 subgroups (Fig. 2). The number of DUF668 genes in each subgroup of *G. hirsutum* and *G. barbadense* was basically twice the number in each subgroup of *G. arboreum* and *G. raimondii*. This was consistent with the results of the previous analysis and conforms to the evolutionary relationship in cotton. The results showed that the DUF668 genes were relatively conserved in evolution in cotton. Although the third subgroup had relatively few members, they were retained during evolution in cotton [22–25], which indicated that they may play an important role in biological processes.

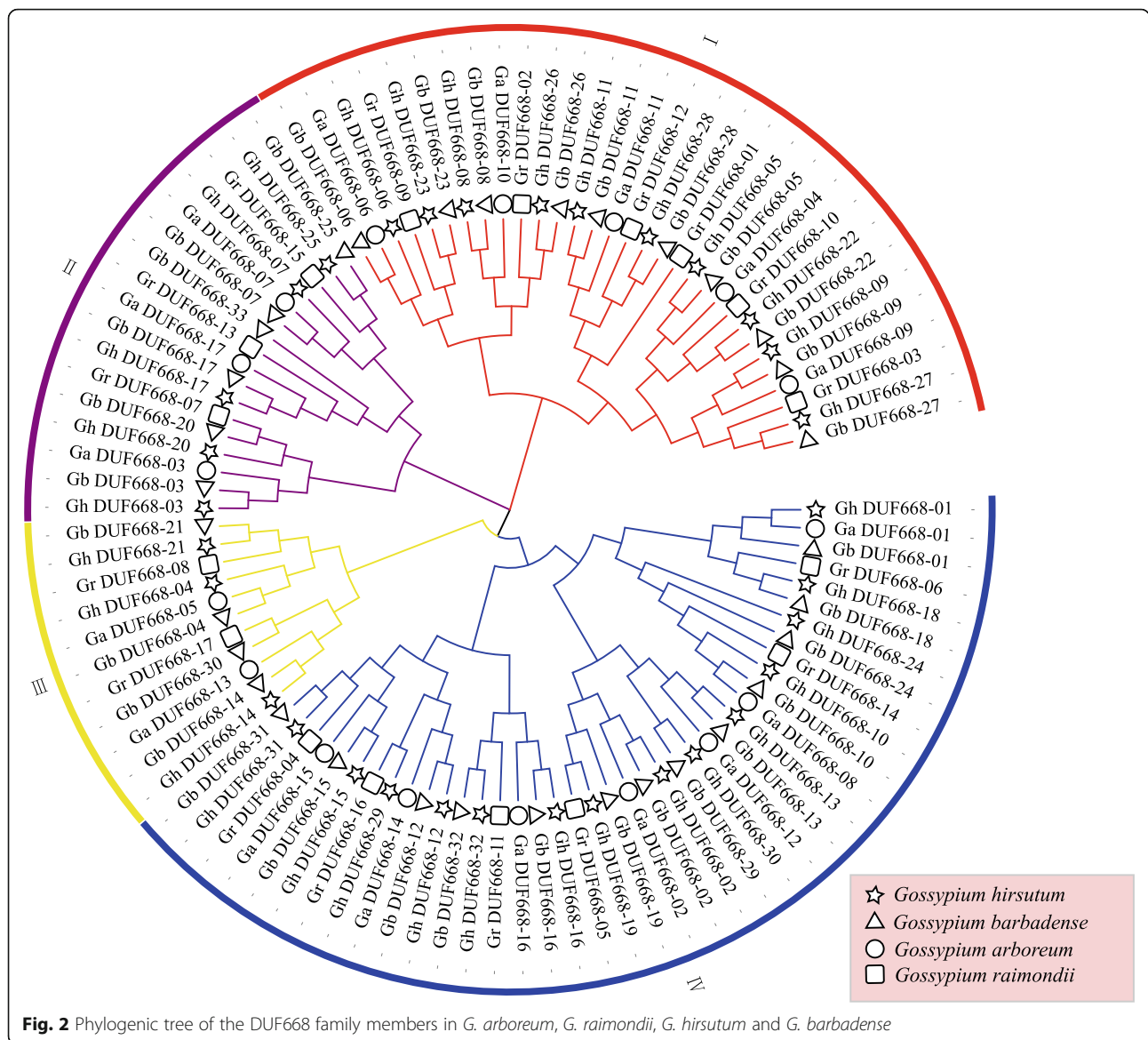
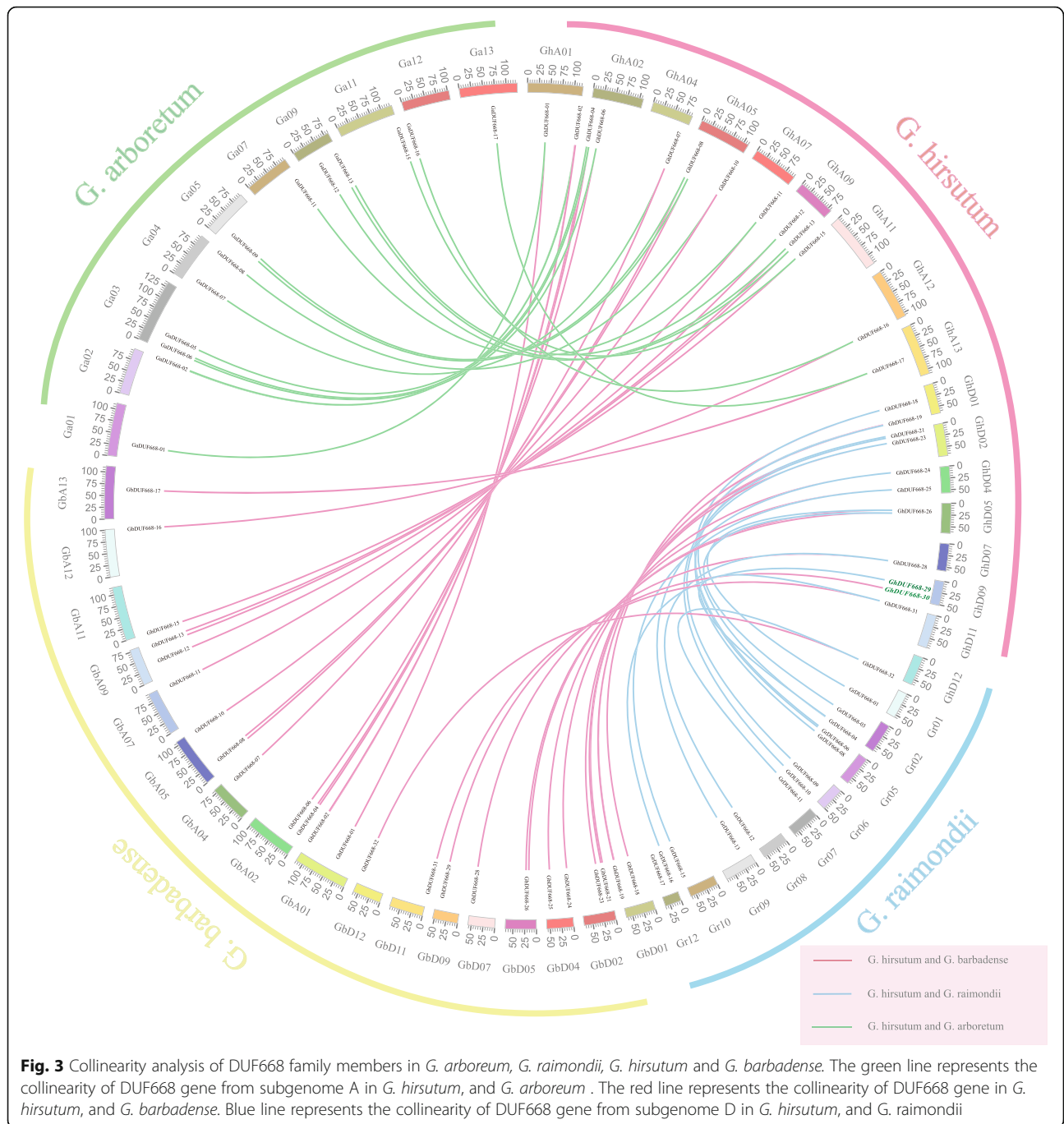


Fig. 2 Phylogenetic tree of the DUF668 family members in *G. arboreum*, *G. raimondii*, *G. hirsutum* and *G. barbadense*

According to the number of genes, chromosome location and phylogenetic tree analysis, DUF668 was predicted to be relatively conserved in cotton. To study the evolutionary relationship of DUF668, we selected *G. hirsutum* as the core and constructed the collinearity relationship in *G. hirsutum* related to other cotton species (Fig. 3). We found that 13 sequences for DUF668 family genes from the subgenome A in *G. hirsutum* had collinearity with 17 sequences in *G. arboreum* and *G. barbadense*. Except for the *Gh_DUF668-30* gene, one sequence for the DUF668 family genes in the subgenome D in *G. hirsutum* had collinearity with one sequence in *G. raimondii* and *G. hirsutum*. However, 11 sequences in *G. barbadense* and 13 sequences in *G. raimondii* had collinearity with 15 and 14 sequences in *G. hirsutum*, respectively. This was basically consistent with the

analytical results of the DUF668 family genes in the A subgroup. Surprisingly, except for *Gh_DUF668-29* and *Gh_DUF668-30*, each sequence of DUF668 family genes in either subgenome A or D in *G. hirsutum* corresponded to only one sequence in *G. arboreum* and *G. barbadense*. This shows that the DUF668 family genes may have been lost during evolution in *G. hirsutum*; later, they were duplicated due to functional requirements, making them consistent with the number in *G. arboreum*. This illustrated the complexity of DUF668 family gene functions.

In order to further study the evolutionary relationship of DUF668 gene family in cotton. Protein sequences of duf668 gene family from *Arabidopsis thaliana*, rice and four different cotton subspecies were selected to construct an evolutionary tree and 10 different conserved

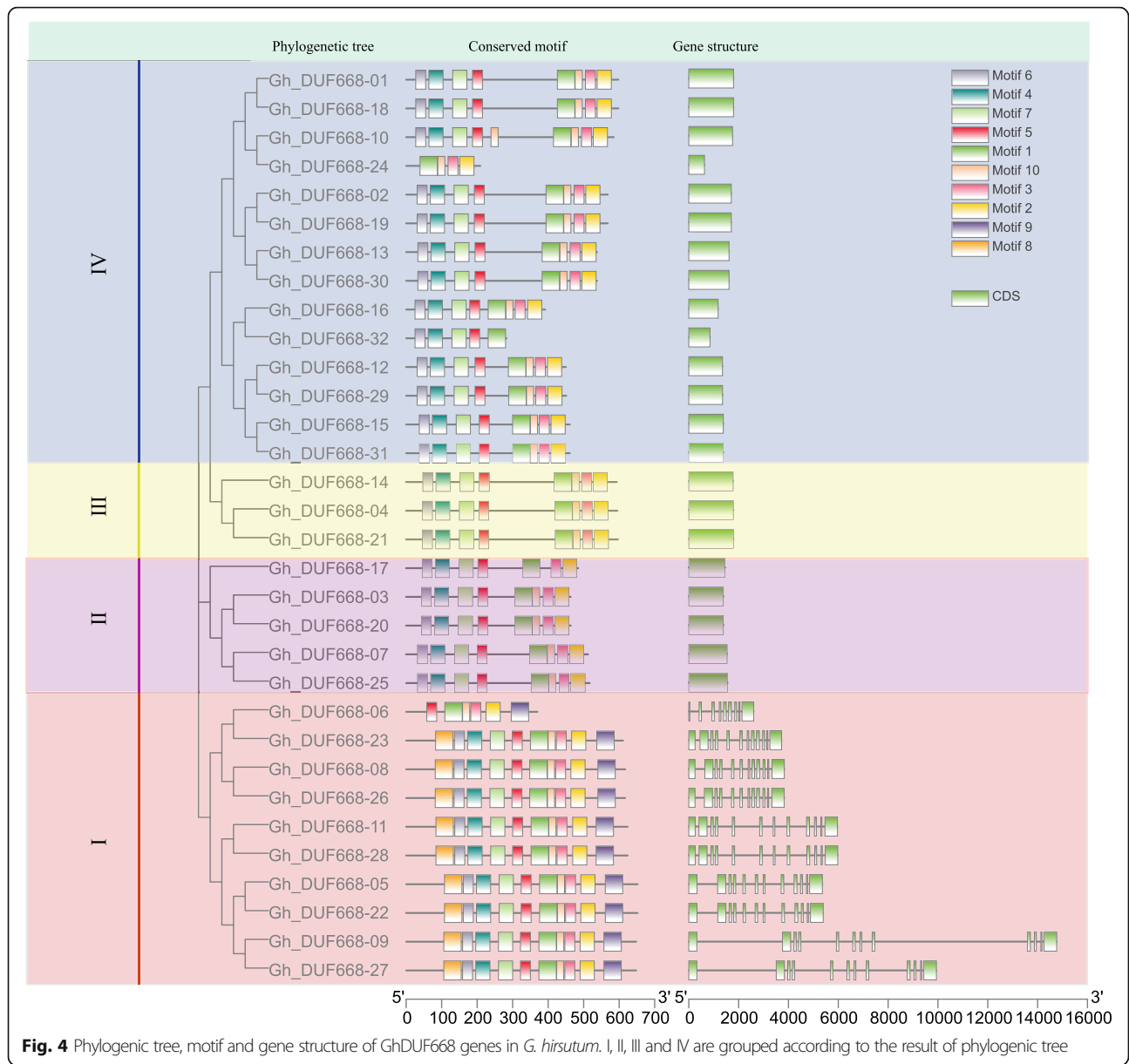


motifs (Figure S2) were identified [2]. The evolutionary tree showed that it could be divided into four categories, which was consistent with cotton. Motif 3, 5.6.7.10 are the most common, which are found in all sequences. Motif 1 and 9 are specific elements of the fourth branch, motif 2 and 4 are specific sequences in addition to the fourth branch. In conclusion, the motifs of DUF668 gene are consistent with their phylogenetic relationships. This indicates that DUF668 gene family has internal

differentiation in the process of evolution, which may further lead to functional differentiation.

Phylogenetic tree, motif and gene structure of the DUF6688 genes in *G. hirsutum*

The phylogenetic tree, gene structure and motif were analyzed according to the full-length coding sequence (CDS) and protein sequence of the GhDUF668 genes (Fig. 4, Fig. S3). Except for *GhDUF668-06*, *24* and *32*,



the rest of the members had the same motif (1, 2, 3, 4, 5, 6, 7, 10), indicating that the same family members had similar functions. Besides the first subgroup, one exon and four identical motifs (1, 2, 3, 10) were observed in other subgroups, whereas introns were not contained. However, the length between the exons was different. Except for *GhDUF668-06*, which contained 6 motifs (1, 2, 3, 5, 9, 10) and 6 exons, the first broad group contained 10 motifs and 12 exons. The difference between the structures of *GhDUF668-06*, 24 and 32 in the same group might be due to changes in the function of the gene or errors in genome annotation. Further study is required. A motif is a structural component with a specific spatial conformation and function in a protein

molecule, which is a subunit of a structural domain and connects with a specific function. This result suggests that the first broad group might have changed its gene structure during the evolutionary process and might have a more important function in cotton growth and development than originally thought.

Cis-acting element analysis of the DUF668 gene in *G. hirsutum*

The 2000 bp promoter region upstream of the GhDUF668 genes was extensively analyzed. Various cis-acting elements were found in defense mechanisms, stress responses, salicylic acid, ABA, gibberellin, auxin, jasmonic acid, light responses, drought induction, MYB

binding sites for flavonoid synthesis, and responses to low temperature, which are related to plant hormones and environmental stress (Fig. 5, Table S3). Previous grouping results showed that the first group contained more cis-acting elements than the other groups, indicating that the first subgroup might have a more important function under adverse stress conditions in cotton growth and development. Each GhDUF668 promoter contained different numbers and types of cis-acting elements, indicating that they might participate in different biotic and abiotic stress responses through the different signaling pathways.

Tissue-specific expression analysis of DUF668 genes in *G. hirsutum*

Gene expression patterns are usually related to gene functions. Our analysis of the expression patterns of the GhDUF668 gene in roots, stems, leaves, pistils, stamens, calyxes, petals and receptacles in cotton showed that most of the selected 32 GhDUF668 genes had tissue expression specificity (Fig. 6) that can be divided into 3 expression patterns. Seven genes (*GhDUF668-01, 18, 05, 22, 24, 09, 27*) could be divided into the first expression pattern, and expression was mainly expressed in the pistil, in roots, stems, and receptacles with the lowest expression in petals and stamens. Nine genes (*GhDUF668-02, 08, 10, 11, 15, 19, 26, 28, 31*) could be divided into the second expression pattern, With most of them expressed in the stem. The rest were classified as the third expression pattern, and the expression level of the eight tissues in this pattern was

low. The gene expression of the GhDUF668 genes was specific and contained more complex functions.

Expression analysis of the DUF668 gene in *G. hirsutum* in response to stress

The expression analysis of the GhDUF668 genes after cold, heat, drought and salt treatment showed that the expression patterns of the GhDUF668 genes can be divided into three categories (Fig. 7). After cold treatment, the expression of nine genes (*GhDUF668-01, -02, -05, -09, -18, -19, -22, -24, -27*) from the C-II and C-III categories significantly changed (Fig. 7a), indicating that the expression of these nine genes could be induced by cold stress. These genes might play a corresponding role in the response to cold stress in *G. hirsutum*. Under heat stress conditions, the expression of 14 genes (*GhDUF668-01, -02, -05, -08, -09, -11, -14, -18, -19, -22, -24, -26, -27, -28*) from the H-II and H-III categories was obviously upregulated and reached the maximum value at 12 h (h) (Fig. 7b), indicating that these genes might play a role in heat resistance in *G. hirsutum*. Polyethylene glycol (PEG) was used to simulate drought stress. The expression of 16 genes (*GhDUF668-01, -02, -05, -08, -09, -10, -11, -14, -18, -19, -22, -23, -24, -26, -27, -28*) was upregulated and reached a maximum at 12 h (Fig. 7c), indicating that these 16 genes might play a role in drought resistance in cotton. After salt stress treatment, the expression of 15 genes (*GhDUF668-01, -02, -05, -08, -09, -10, -11, -14, -18, -19, -22, -24, -26, -27, -28*) from the S-II and S-III categories was upregulated and reached a maximum at 12 h (Fig. 7d), indicating that these 15 genes might play a role in salt tolerance in

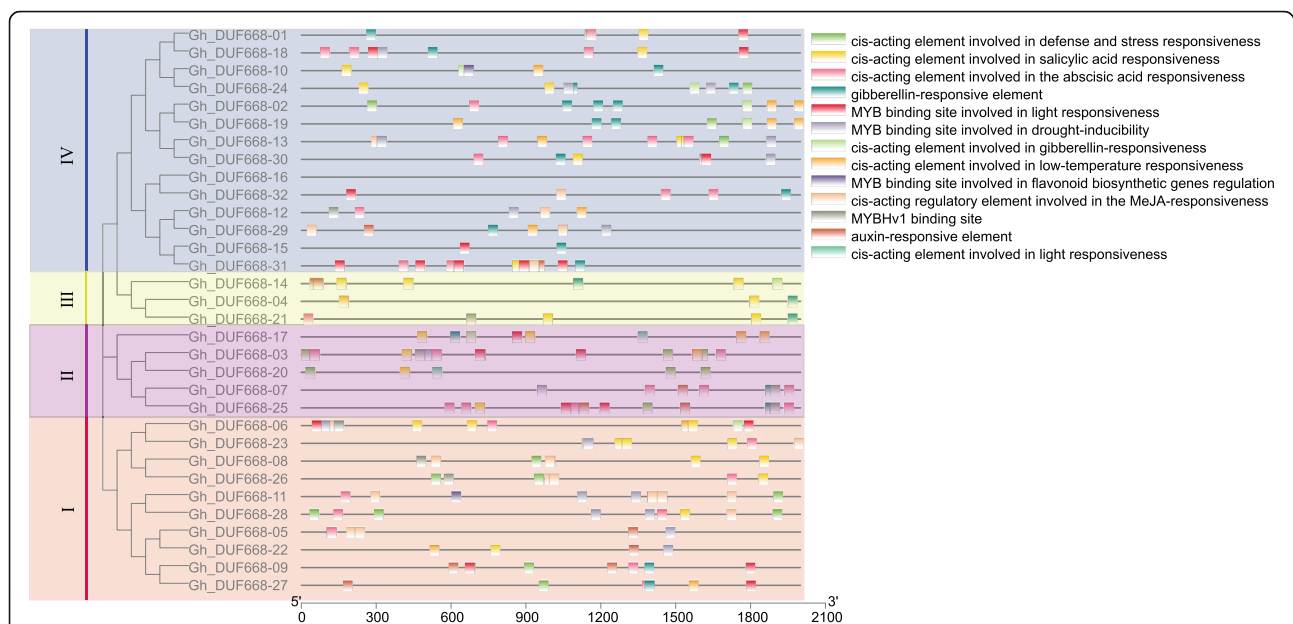


Fig. 5 Cis-acting element analysis of DUF668 family members in *G. hirsutum*. I, II, III and IV are grouped according to the result of phylogenetic tree

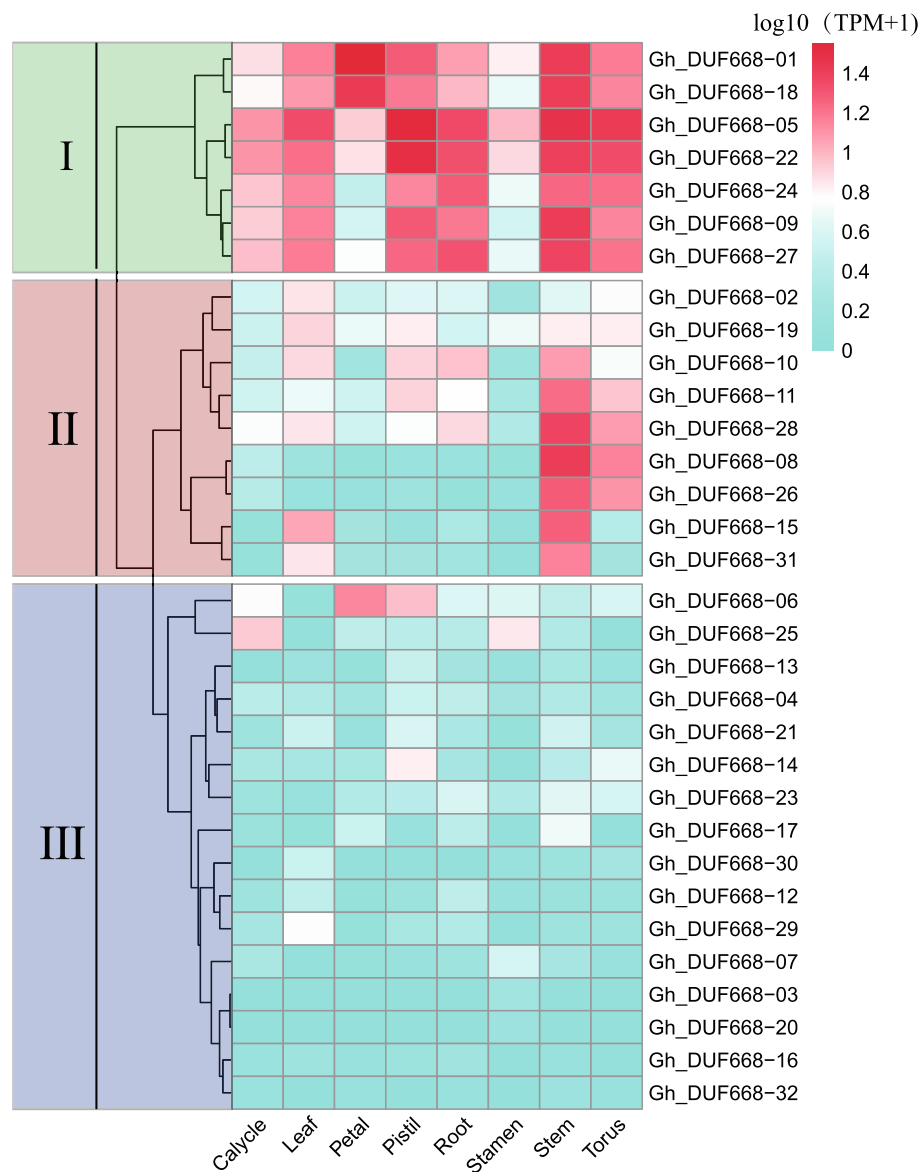


Fig. 6 Tissue-specific expression analysis of DUF668 genes from *G. hirsutum* based on transcriptome data. I, II and III are grouped and clustered according to the expression pattern

cotton. In summary, nine genes (*GhDUF668-01*, *-02*, *-05*, *-09*, *-18*, *-19*, *-22*, *-24*, *-27*) were induced by various abiotic stresses, implying that they might be affected by non-biological stress in cotton. Although most of the genes had obvious expression changes and reached their maximum at 12 h, RNA-seq measurement was performed at only 12 h. We speculated that these genes might continue to increase their expression after 12 h under abiotic stress conditions in *G. hirsutum*. In summary, the expression levels of approximately half of the GhDUF668 genes significantly changed with higher tissue-specific expression levels under abiotic stress conditions. This result indicated that the GhDUF668

genes might play a role in the abiotic stress response of *G. hirsutum*. However, further verification is needed.

Expression analysis of the DUF668 gene in *G. hirsutum* in response to Verticillium wilt stress

Cotton production is restricted by *V. dahliae*, which has a serious impact and causes great economic losses every year. This study was based on the transcriptomic data of cotton roots after inoculation with *V. dahliae*. The results showed (Fig. 8) that these GhDUF668 genes could be clustered into three expression patterns within 0 ~ 120 h. The expression of 20 genes (*GhDUF668-01*, *-02*,

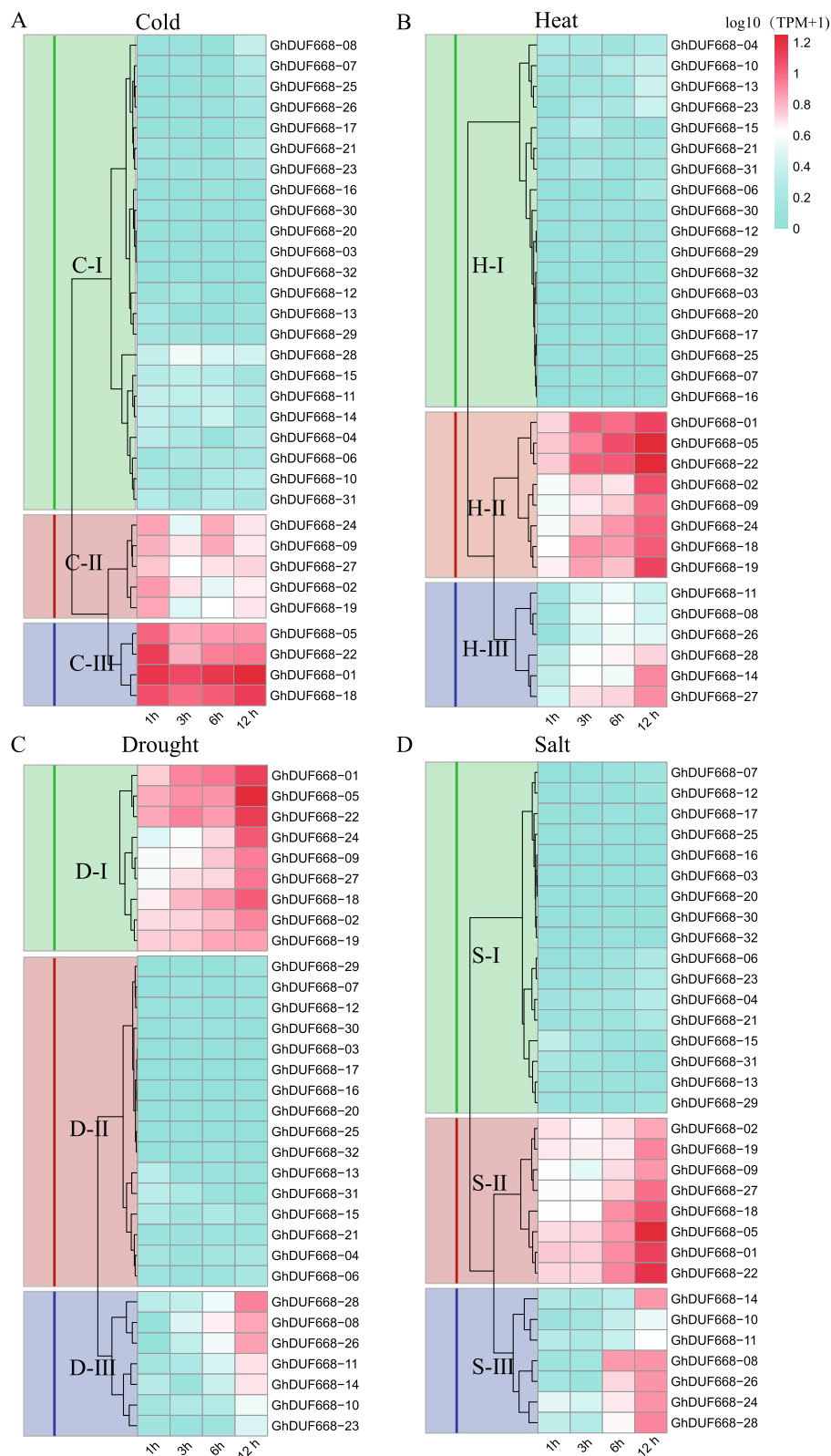


Fig. 7 Expression analysis of the DUF668 genes from *G. hirsutum* under cold (a), heat (b), drought (c) and salt (d) stress based on transcriptome data. I, II and III are grouped and clustered according to the expression pattern. C, H, D, S represent cold, heat, drought and salt, respectively

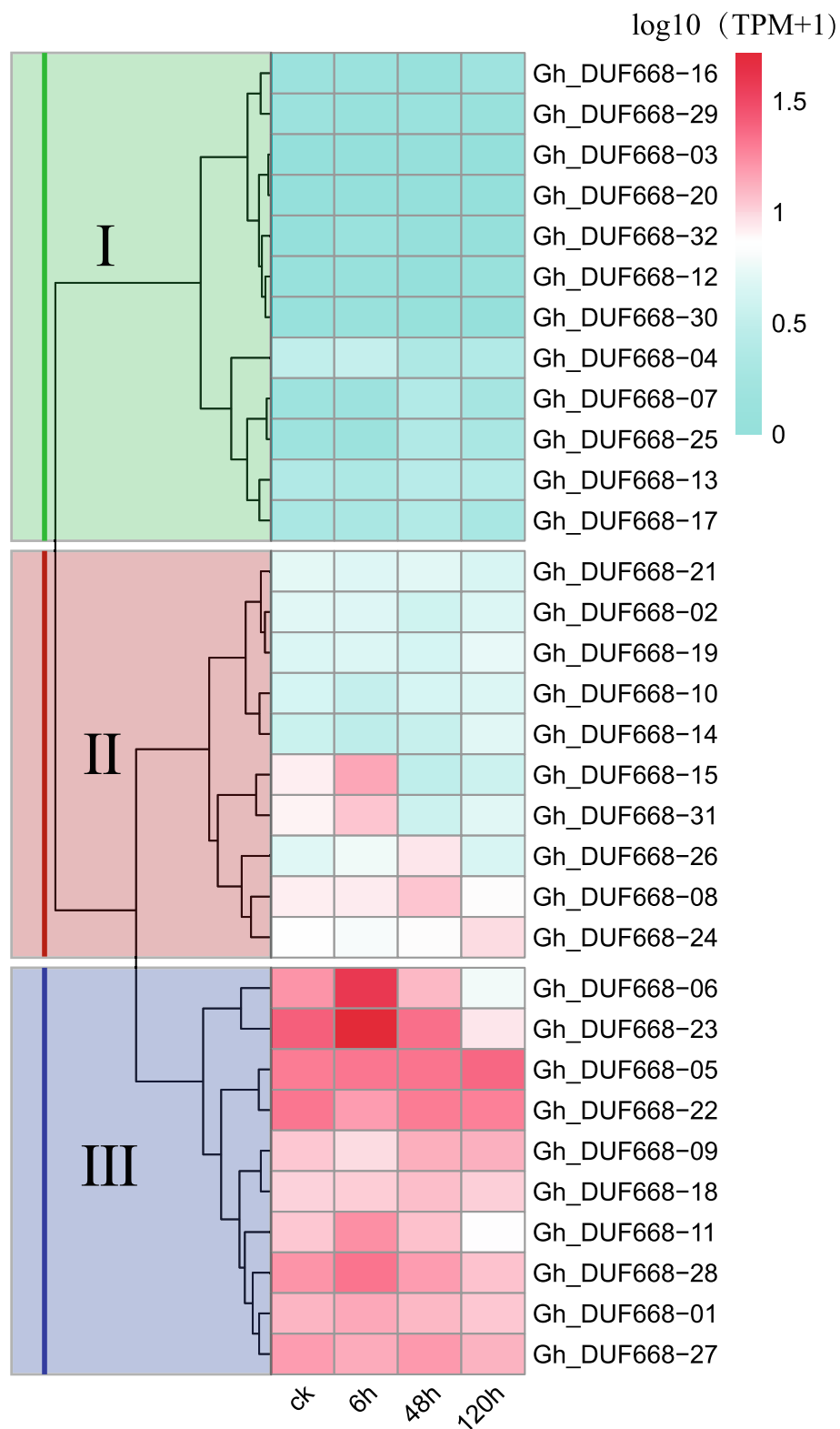


Fig. 8 Expression analysis of the DUF668 genes from *G. hirsutum* under *V. dahliae* stress based on transcriptome data. I, II and III are grouped and clustered according to the expression pattern

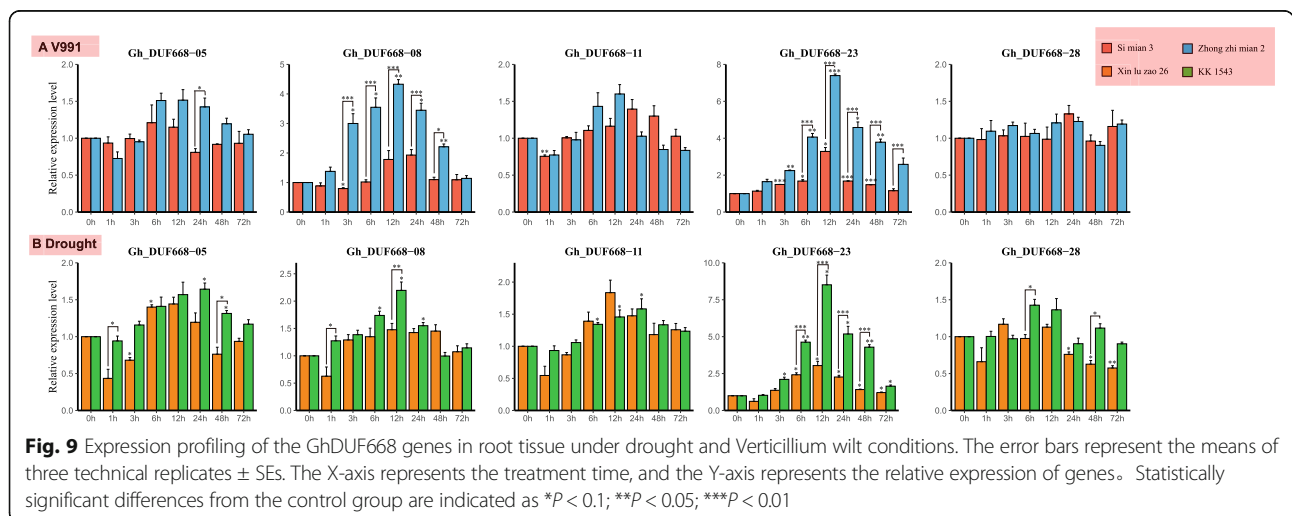
- 05, - 06, - 08, - 09, - 10, - 11, - 14, - 15, - 18, - 19, - 21, - 22, - 23, - 24, - 26, - 27, - 28 and - 31) from the second and third expression patterns changed. Among them, the expression levels of six genes (*GhDUF668-06*, - 11, - 15, - 23, - 28, and - 31) reached a maximum at 6 h and then continued to decline, indicating that these six genes might be involved in the response to *V. dahliae* treatment in the early stage of cotton. The expression levels of *GhDUF668-08* and *GhDUF668-26* reached a maximum at 48 h and then decreased, indicating that these two genes may play a role in the middle and late stages in response to *V. dahliae* infection in cotton.

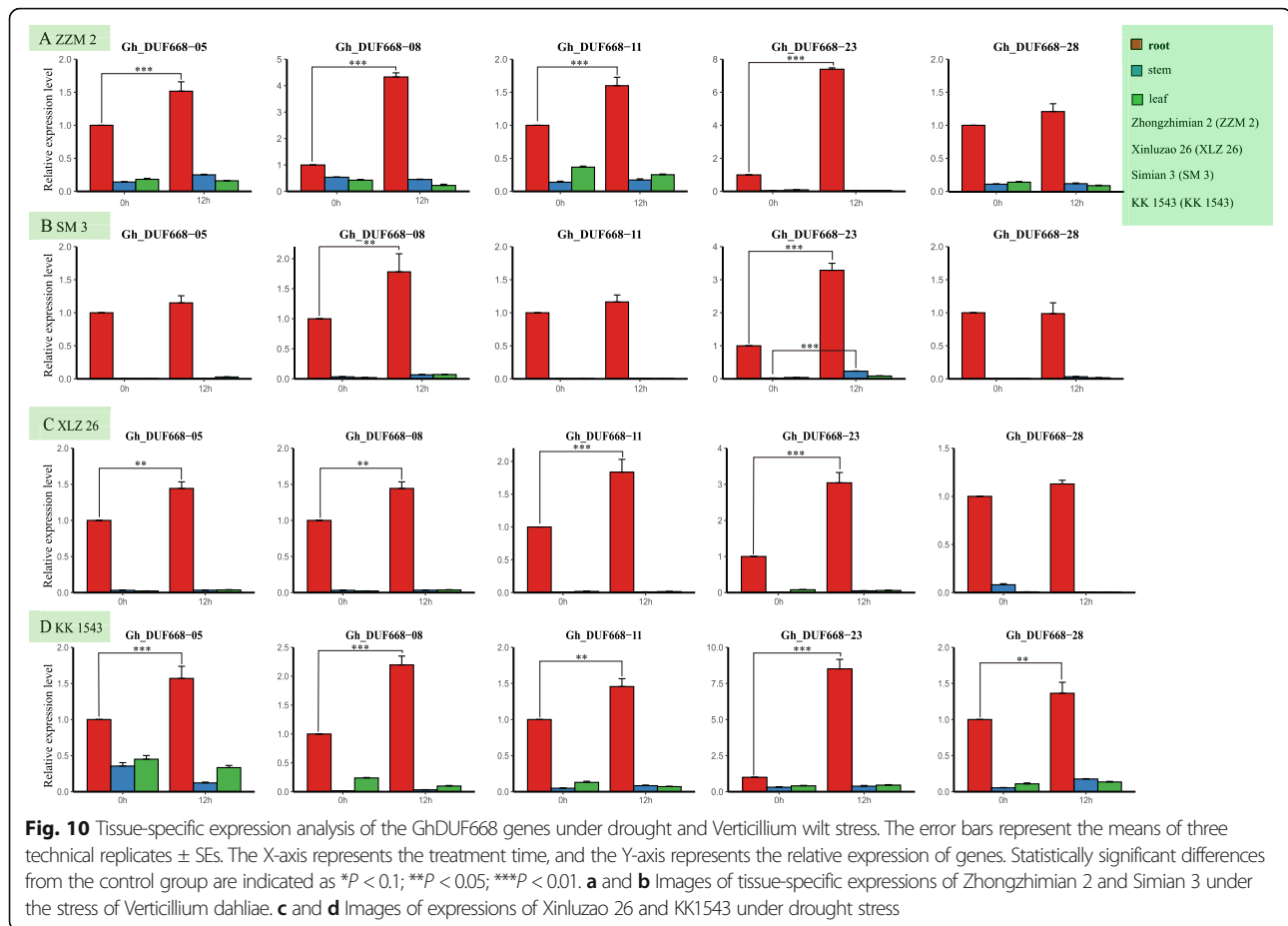
Response of the DUF668 gene in *G. hirsutum* to drought and Verticillium wilt stress by qRT-PCR

Adverse stresses can cause transcriptome reprogramming events. Our previous analysis showed that most of the expression of the GhDUF668 genes undergoes significant changes under stress treatment conditions. Based on the transcriptome expression profiles and promoter cis-acting elements, we speculated that five genes (*GhDUF668-05*, - 08, - 11, - 23 and - 28) from the GhDUF668 genes might be involved in resistance to stress conditions. We selected KK 1543 (drought-resistant), Xinluzao 26 (drought-susceptible), Zhongzhimian 2 (disease-resistant) [28] and Simian 3 (disease-susceptible) (unpublished) to determine whether these genes were involved in the response to adverse stress. qRT-PCR was used to detect the transcription levels of these five GhDUF668 genes in roots under drought and Verticillium wilt treatment at the seedling stage. Compared with that at 0 h, the expression of these 5 genes was significantly different after the roots were stressed (Fig. 9). The transcription levels of these five genes were significantly induced after adverse stress treatment.

After the roots of disease-resistant and susceptible materials were inoculated with *V. dahliae*, the RNA transcription levels of selected genes were significantly induced at different periods in the two materials, while except for Gh-DUF668-28. Thus, the expression levels were significantly increased, suggesting that these genes might participate in the process of responding to the invasion of V991 in cotton. Among them, *GhDUF668-08* and *GhDUF668-23* reached their maximum values at 12 h in disease-resistant materials. The maximum expression of GhDUF668-08 was observed at 24 h in susceptible materials, and expression was lower in the susceptible materials than in the disease-resistant materials. Although *GhDUF668-08* and *GhDUF668-23* were significantly induced in both materials, the expression level in disease-resistant materials was dramatically higher than that in susceptible materials. Minor changes in the expression levels of the other three genes were observed. In summary, these five genes might play a role in the process of responding to the invasion of *V. dahliae* in cotton. Among them, *GhDUF668-08* and *GhDUF668-23* might play leading roles in disease resistance.

After PEG simulated drought treatment, the transcription levels of all selected genes were significantly induced at different periods in the two materials, and the expression levels were significantly increased, suggesting that these genes may be involved in the response to drought conditions in *G. hirsutum*. *GhDUF668-08* and *GhDUF668-23* were significantly induced at 6 h, and their expression increased sharply, reaching a maximum at 12 h. The changing expression trends of *GhDUF668-08* and *GhDUF668-23* under drought treatment were consistent with those of inoculation with V991. However, the expression level of *GhDUF668-08* in *G. hirsutum* under drought treatment was not significantly





different from that in *G. hirsutum* inoculated with *V. dahlia*. In summary, these 5 genes might have a role in the drought stress response in cotton. Furthermore, *GhDUF668-23* might play a leading role in this process.

The same expression pattern was observed among five genes (*GhDUF668-05*, *-08*, *-11*, *-23* and *-28*) in different materials under biotic stress and abiotic stress. Compared with expression at 0 h, expression of 5 genes at other timepoints changed significantly. This implied that the expression of these genes was regulated by adverse stresses and might play a certain role in the process of responding to adverse stress in *G. hirsutum*. Among them, the expression levels of both *GhDUF668-08* and *GhDUF668-23* increased sharply at 6 h and reached a maximum at 12 h with drought stress and *V. dahlia* inoculation treatment, respectively. The expression levels of *GhDUF668-08* and *GhDUF668-23* were significantly upregulated among the five genes in the same period. This further shows that *GhDUF668-08* and *GhDUF668-23* might play certain roles in the response to biotic and abiotic stress in cotton.

Tissue-specific expression analysis of the DUF668 gene in *G. hirsutum*

We proved that the expression of *GhDUF668-08* and *GhDUF668-23* changed dramatically under different stress conditions, and significant differences in their expression were observed among different materials. Moreover, due to the specific parts of gene expression, the molecular biological functions were different. The first part of the plants to sense the response to drought stress and pathogen invasion is the root. For this reason, we selected the periods that could strongly verify the tissue-specific expression of these five candidate genes in the four materials during treatment. We found that the expression in roots was significantly higher than that in stems and leaves (Fig. 10). Interestingly, expression of these five genes changed significantly in the roots before and after different stresses, but severe changes occurred in the stems and leaves. Among them, the *GhDUF668-08* and *GhDUF668-23* genes changed significantly in the roots. This result showed that the increased expression of the GhDUF668 genes in the roots was likely to enhance cotton resistance. These results further illustrated

the certain role of the GhDUF668 genes in stress resistance.

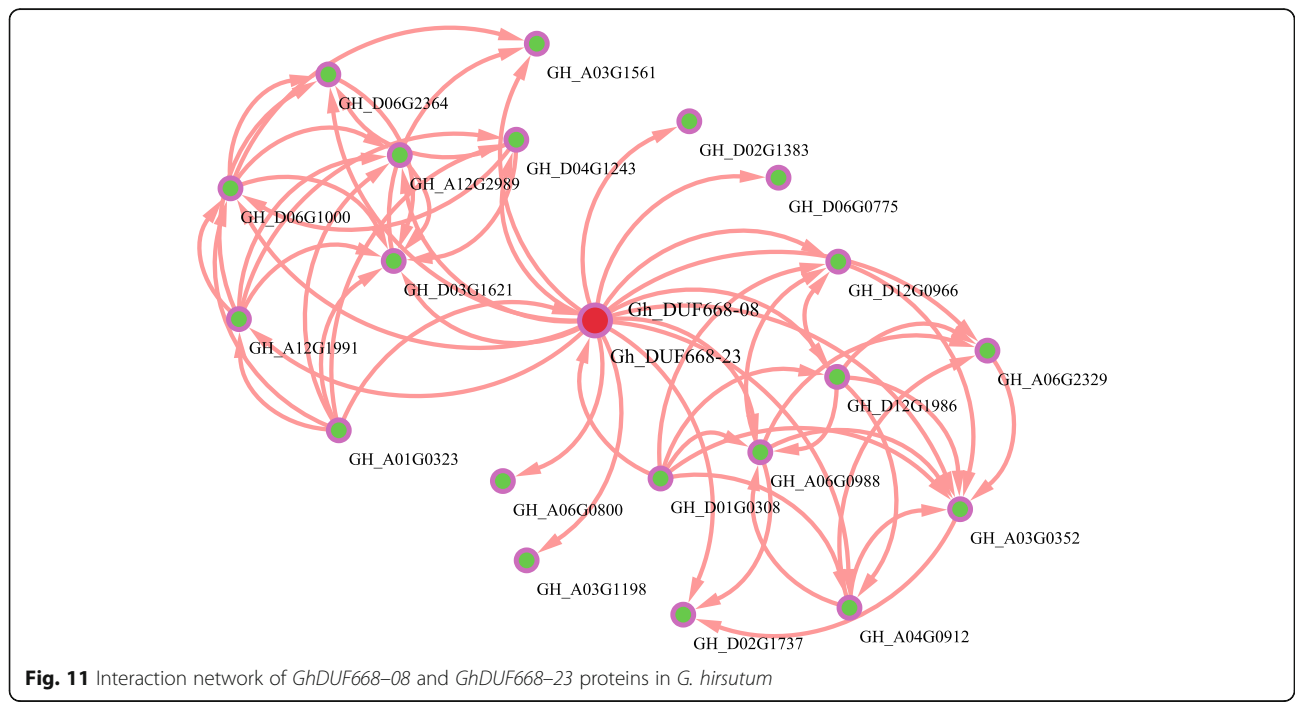
Analysis of protein interaction networks

As we all know, rarely single protein can directly participate in plant stress response. Most of plant physiological processes are completed by protein-protein interaction. For better understanding the molecular mechanism of GhDUF668-08 and GhDUF668-23 genes The interaction network between GhDUF668-08 and GhDUF668-23 genes proteins and other *Gossypium hirsutum* proteins was constructed (Fig. 11), based on the interaction network of Arabidopsis homologous genes (Figure S4). The interaction between GhDUF668-08, GhDUF668-23 proteins with 20 *Gossypium hirsutum* proteins was found. Among them, GH_A03G1561 and GH_D01G0308 has played a potential role in response to stress in *G. hirsutum*. Only GH_A01G0323 and GH_D01G0308 are located upstream of GhDUF668-08 and GhDUF668-23, which indicates that GH_A01G0323 and GH_D01G0308 may interact with the DUF3475 domain of GhDUF668-08 and GhDUF668-23. However, more genes are combined with the downstream DUF668 domain of GhDUF668-08 and GhDUF668-23, which further indicates the complex function of DUF668 gene family and the potential role of GhDUF668-08 and GhDUF668-23 in response to stress in *G. hirsutum*. Our results lay a molecular foundation to further investigate the function and molecular mechanism of these genes in stress tolerance.

Discussion

Many studies have shown that genome-wide identification and expression analysis could help researchers understand the origin, diversity and biological functions of these DUF gene families [2, 5, 8, 9, 31]. In recent years, with the improvement of sequencing technology and the decrease in sequencing cost, cotton genome sequencing has been continuously improved and updated, laying the foundation for studying gene functions at the comprehensive genome level [22–25]. This study further deepens the knowledge of cotton genomics and genetics, providing the possibility to discover members of the DUF668 gene family in cotton and their phylogenetic relationships. At present, research on cotton DUF668 family genes has not been reported.

Our research shows that more DUF668 family genes exist in cotton than in monocot and dicot plants [2]. This might contribute to the double duplication of the cotton genome, or it might be that more genes were annotated in cotton than in other species [24, 25, 32, 33]. The number of genes, phylogenetic tree and collinearity showed (Figs. 2 and 3) that the cotton DUF668 gene family was conserved during the evolution of cotton, which was consistent with the evolutionary relationships between cotton species [34–37]. In cotton, a multitude of gene family members with high sequence similarity have been observed to cluster on chromosomes as paralogous pairs [38–40]. However, the GhDUF668 genes did not do so, proving that this gene family has not undergone gene amplification events, including tandem



and fragment duplication, during evolution. Protein structure and gene structure are closely related to gene function. The numbers of intron/exon might represent splicing variants and were used to classify genes [41, 42]. According to the distribution and quantity of gene structures and motifs analysis, DUF668 members were divided into 2 broad groups in *G. hirsutum*, which was consistent with research in rice [2]. The number and structure of genes in each broad group were significantly different (Fig. 4). Moreover, the number of genes from the two broad groups was basically the same in rice and some Poaceae [2]. Although the second broad group contains the same number of CDS and motif except for Gh_DUF668-06, from the perspective of the full length of genes, the number and length of similar CDS have different gene lengths, indicating that the intron length has a larger difference. Surprisingly, we found that in addition to *GhDUF668-06* and *GhDUF668-24*, the DUF668 gene family contains not only the DUF668 domain but also the DUF3475 domain in *G. arboreum*, *G. raimondii* and *G. hirsutum*. However, the DUF3475 domain in plants has not been reported. We clearly found that motif is located in the two domains respectively (Fig. 4, Fig. S1). Our results showed that the GhDUF668 genes might belong to the same family as GhDUF3475 in cotton. DUF668 contains more complex functions and may also be related to the DUF3475 domain. Therefore, we speculated that the DUF3475 domain might bind to upstream transcription factors, and certain functions were performed by regulating downstream genes through DUF668. These results indicate that this gene family plays a more important role in the growth and development of cotton.

Expression patterns of genes can provide important clues for characterizing gene function, which is thought to be related to promoter region differentiation. Through the analysis of promoters, tissue-specific expression and expression profiles (Figs. 5, 6, 7, 8), we found that *Gh_DUF668-05*, *Gh_DUF668-08*, *Gh_DUF668-11*, *Gh_DUF668-23* and *Gh_DUF668-28* contained more homeopathic action elements, and higher expression was observed under adverse stress conditions. We speculated that these five genes might participate in resistance to stresses in cotton development. However, we were surprised to find that these five genes all came from the second broad group. Although the second broad group contains a relatively small number of genes, they have been retained during the evolution of cotton, implying that they may play important roles in biological processes. Regardless of whether plants are subjected to biological or abiotic stress, their initial sensing parts are basically at the roots [43–45]. Moreover, analysis of the gene expression in different tissues and the molecular

biological functions were performed differentially [46, 47]. The *Gh_DUF668-08* and *Gh_DUF668-23* expression levels both increased sharply at 6 h and reached a maximum at 12 h under biotic and abiotic stresses. An obvious difference was found between these two genetically resistant materials in the same treatment period. Moreover, we found that expression in the roots was significantly higher than that in the stems and leaves. However, minor changes in the stems and leaves were observed. Among the genes, *Gh_DUF668-08* and *Gh_DUF668-23* had the most obvious changes in the roots.

Many studies have shown that salicylic acid (SA), methyl jasmonate (MeJA) and ABA are apparently related to plant resistance [48–52]. A large number of cis-acting elements were contained in *Gh_DUF668-08* and *Gh_DUF668-23*, which are involved in defense mechanisms, responses to stresses, and SA, ABA, MeJA and MYB binding sites (Fig. 5). Homology analysis showed that *Gh_DUF668-08* and *Gh_DUF668-23* have the highest homology with *Arabidopsis AT1G30755*, which is highly related to growth, development and defense in *Arabidopsis* [53]. In a study conducted with flg22 treatment in *Arabidopsis*, the expression levels of 35 genes that contain the DUF domain were obviously changed, and DUF668 members were included [54]. The results indicate that *Gh_DUF668-08* and *Gh_DUF668-23* play leading roles in the response to biotic and abiotic stresses in cotton.

Thus, cotton tissues and expression profiles under various stress conditions were analyzed using qRT-PCR to detect the GhDUF668 genes. It was shown that most members of this family were very important in the process of resistance to one or more stresses in *G. hirsutum*. Moreover, we speculated that the apparent changes in GhDUF668 gene expression in the roots were modulated by adverse stresses, which caused the enhancement of cotton resistance. These results laid the foundation for us to further verify its function and the molecular mechanism of stress resistance.

Conclusions

In this study, cotton DUF668 genes were identified in the whole genome. According to the number of genes and phylogenetic tree and collinearity analyses, the cotton DUF668 family genes could be divided into 4 subgroups. Thirty-two DUF668 genes were observed in the *G. hirsutum* genome, and their structural characteristics and expression patterns were analyzed. We found that the genes from the GhDUF668 gene family could be divided into two broad groups. *GhDUF668-08* and *GhDUF668-23*, which might play leading roles in the response of biotic and abiotic stress in cotton.

This study groundbreakingly conducted a systematic analysis with the cotton DUF668 gene family and provided a new understanding of cotton resistance to stress, laying the foundation for the in-depth functional analysis and breeding application of these genes.

Methods

Plant material

Based on the previous research results of our research team, expression analysis of KK1543 (drought resistance, Xinjiang Bazhou Agricultural Institute), Xinluzao 26 (drought susceptibility, Xinjiang Academy of Agricultural Sciences), Zhongzhimian 2 (disease resistance, The Institute of Plant Protection, Chinese Academy of Agricultural Sciences) and Simian 3 (disease susceptibility, Jiangsu Academy of Agricultural Sciences) materials was conducted under drought and Verticillium wilt (V991) inoculation. The seeds of KK1543 and Xinluzao 26 were germinated at 28 °C under a 16 h light/8 h dark cycle and then transplanted into normal hydroponic solution. Hoagland nutrient solution was applied every 2 days. KK1543 and Xinluzao 26 were subjected to drought treatment with 15% PEG6000 at the two-leaf stage. Zhongzhimian 2 and Si Mian 3 were planted in soil pots [28]. The middle and lower parts of the pot were cut with a small knife in the two-leaf stage to ensure that the roots were damaged, then *V. dahliae* 991, the spore suspension (1×10^7 spores/ml) was watered along the roots of cotton seedlings to ensure they were successfully inoculated. The root tissues of the two treatments were collected at 0, 1, 3, 6, 12, 24, 48, and 72 h.

Identification and bioinformatics analysis of the cotton DUF668 gene

This study used the latest published genome sequence as a reference. The relevant genomic and proteomic data in *G. arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense* were downloaded from COTTONGEN (<http://www.cottongen.org/>) [55]. Using the hidden Markov model (PF05003) of the DUF668 gene domain, the protein sequence was identified in cotton through the hidden Markov model (HMM) from HMMER3.2.1 software (<http://hmmer.org/>) [4, 56]. After the redundancies were removed, domain verification for candidate genes was conducted in the NCBI-CDD (<https://www.ncbi.nlm.nih.gov/cdd/>) database. Therefore, cotton DUF668 gene family members were further confirmed [57]. The number of amino acid residues, relative molecular mass, and theoretical isoelectric point of the GhDUF668 protein were calculated by ExPASy software (<http://cn.expasy.org/tools>) [58]. Subcellular localization of the GhDUF668 protein was predicted by EuLoc online software (<http://euoloc.mbc.nctu.edu.tw/>) [59].

Phylogeny and collinearity analysis of the cotton DUF668 gene family

DUF668 protein sequences identified from *G. arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense* were subjected to alignment by Clustal W with the default settings in MEGA 7 software. A phylogenetic tree was constructed based on the results of the sequence alignment by using neighbor-joining (NJ), and the bootstrap value setting was 1000 [60]. The phylogenetic tree was edited and displayed with the online tool Evolview (<https://evolgenius.info/>) [61]. The chromosomal positions and collinearity of the DUF668 genes in cotton were visualized through Circos (<http://circos.ca/>).

Chromosome location, gene structure and motif analysis of the DUF668 gene family in *G. hirsutum*

The chromosome location information of DUF668 gene family members was extracted from the *G. hirsutum* genome annotation file, and the chromosome location map of the GhDUF668 gene was redrawn with Mapchart software [62]. The phylogenetic tree of the GhDUF668 genes was constructed, and the nwk file was obtained through MEGA7 software. Motif analysis was performed using the MEME program (the maximum number of motifs was set to 10, [63]). The xml file obtained after running the MEME program, the nwk file obtained from the evolutionary tree, and the gff file obtained from the gene structure were processed and visualized with TBtools software [64].

Analysis of upstream cis-acting elements of the DUF668 gene

The upstream DNA sequences (2000 bp) of the GhDUF668 genes were obtained. Then, the possible cis-acting elements were predicted through the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and visualized with TBtools [65].

RNA-seq analysis

Transcriptome data including tissues (roots, stems, leaves, pistils, stamens, calyxes, petals and receptacles) under adverse stress (cold, heat, drought and salt) conditions (Genome sequencing project accession: PRJNA248163) and *V. dahliae* inoculation (Genome sequencing project accession: SRP118279) of *G. hirsutum* were downloaded from the NCBI Sequence Read Archive (SAR) database. The expression data were analyzed by standard analysis of the transcriptome on the raw data [66–68]. Log₁₀ (TPM + 1) normalization was performed on the expression data. The standardized data were drawn with R-4.0.2 language.

qRT-PCR

According to the cDNA information of the GhDUF668 genes, primers in the specific region were designed at

the 5' end or 3' end of the gene sequence with Primer 5.0 software (Table S4). cDNA extracted from root tissue was used as the template, and the expression of candidate genes was detected with qRT-PCR. The GhUBQ7 gene was chosen as the reference gene for qRT-PCR analysis. Each entire experiment was repeated three times.

qRT-PCR was performed according to a previously reported method [30]. In brief, total RNA was isolated using a total RNA extraction kit (Tiangen, China). Then, 1 µg of total RNA after DNase I digestion was reverse transcribed into cDNA using 5× All-In-One Mastermix (Abm, Canada). Real-time PCR amplification was performed with an ABI 7500 Fast Real-Time PCR system. The cDNA was amplified using a BrightGreen 2× qPCR Mastermix (Abm, Canada) kit. The reaction program was a thermal cycling program at 94 °C for 30 s followed by 40 cycles of 95 °C for 5 s, 57 °C for 5 s, and 72 °C for 34 s. The relative RNA transcript levels of the candidate genes were calculated according to the $2^{-\Delta\Delta CT}$ method [69].

Protein interaction network analysis

A comparison of GhDUF668 gene was performed in the Arabidopsis information resource (<https://www.arabidopsis.org/>) by using homologous protein analysis. The homologous gene of GhDUF 668 in *Arabidopsis thaliana* was screened out and submitted to database of STRING (<https://string-db.org/>). And the relationship of protein-protein interaction network of *Arabidopsis thaliana* was further obtained from the database. Then, the homologous genes of *Gossypium hirsutum* were screened out by a comparison of the interaction genes from Arabidopsis in COTTONGEN (<http://www.cottongen.org/>) [55]. The software Cytoscape-3.8.2 was performed for visualization [70].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07716-w>.

Additional file 1: Figure S1. Distribution of the DUF668 domain in the DUF668 proteins of cotton.

Additional file 2: Table S1. Information on the DUF668 gene family in *G. barbadense*, *G. arboreum* and *G. raimondii*.

Additional file 3: Table S2. cotton DUF668 gene protein sequences.

Additional file 4: Figure S2. Conservative motifs of DUF668 gene family in *Arabidopsis thaliana*, *rice*, *G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii*.

Additional file 5: Figure S3. Conservative motifs of DUF668 gene family in *G. hirsutum*.

Additional file 6: Figure S4. Interaction network of AT1G30755 proteins in Arabidopsis.

Additional file 7: Table S3. cis-acting elements analysis of DUF668 family members in *G. hirsutum*.

Additional file 8: Table S4. All primers used in this study.

Acknowledgments

Not applicable

Authors' contributions

Conceived and designed the experiments: JYZ, QJC and YYQ. Collected public datasets and performed experiments: JYZ, PW, WJG and YLL. Analyzed the data: JYZ, YXW, SWG, YJ and XNS. Wrote the manuscript: JYZ and PW. Revised the manuscript: JYZ, QJC and YYQ. All authors read and approved the final manuscript

Funding

This work was supported by the Major Special Project of Xinjiang Uygur Autonomous Region (2020A01002–2) and New Cultivar Breeding of Transgenic Herbicide Resistance in Cotton (2020ZX08005–005). The funding agencies played no role in the design of the study, data collection or analysis, or interpretation or writing of the manuscript.

Availability of data and materials

The genome databases were downloaded from COTTONGEN (<http://www.cottongen.org/>) and RNA-seq (<https://www.ncbi.nlm.nih.gov/sra/?term=>). The datasets supporting the conclusions of this article are included in the article and its Additional files.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests.

Received: 17 December 2020 Accepted: 14 May 2021

Published online: 27 May 2021

References

- Ohta T. Evolution of gene families. *Gene*. 2000;259(1–2):45–52. [https://doi.org/10.1016/S0378-1119\(00\)00428-5](https://doi.org/10.1016/S0378-1119(00)00428-5).
- Zhong H, et al. Characterization and functional divergence of a novel DUF668 gene family in rice based on comprehensive expression patterns. *Genes (Basel)*. 2019;10(12):980.
- Bateman A, Coggill P, Finn RD. DUFs: families in search of function. *Acta Crystallogr Sect F Struct Biol Cryst Commun*. 2010;66(Pt 10):1148–52. <https://doi.org/10.1107/S1744309110001685>.
- El-Gebali S, et al. The Pfam protein families database in 2019. *Nucleic Acids Res*. 2019;47(D1):D427–32. <https://doi.org/10.1093/nar/gky995>.
- Ganie SA, Pani DR, Mondal TK. Genome-wide analysis of DUF221 domain-containing gene family in *Oryza* species and identification of its salinity stress-responsive members in rice. *PLoS One*. 2017;12(8):e0182469. <https://doi.org/10.1371/journal.pone.0182469>.
- Li HL, et al. The rice OsDUF810 family: OsDUF810.7 may be involved in the tolerance to salt and drought. *Mol Biol*. 2018;52(4):567.
- Chengke L, et al. Overexpression of a new stress-repressive gene OsDSR2 encoding a protein with a DUF966 domain increases salt and simulated drought stress sensitivities and reduces ABA sensitivity in rice. *Plant Cell Rep*. 2014;33(2):323.
- Lihua L, et al. Genome-wide identification and analyses of the rice OsDUF936 family. *Biotechnol Biotechnol Equip*. 2018;32(2):1–7.
- Wang L, et al. Characterization of a novel DUF1618 gene family in rice. *J Integr Plant Biol*. 2014;56(2):151.
- Xi C, et al. Corrigendum: The complex becomes more complex: protein-protein interactions of SnRK1 with DUF581 family proteins provide a framework for cell- and stimulus type-specific SnRK1 signaling in plants. *Front Plant Sci*. 2014;5:693.
- Börnke F. Corrigendum: The complex becomes more complex: protein-protein interactions of DUF724 gene family in *Arabidopsis thaliana*. *Plant Mol Biol*. 2010;72(1–2):61.
- Karras GJ, Kustatscher G, Buhecha HR, Allen MD, Pugieux C, Sait F, et al. The macro domain is an ADP-ribose binding module. *EMBO J*. 2005;24(11):1911–20. <https://doi.org/10.1038/sj.emboj.7600664>.

13. Dlakic M. DUF283 domain of dicer proteins has a double-stranded RNA-binding fold. *Bioinformatics*. 2006;22(22):2711–4. <https://doi.org/10.1093/bioinformatics/btl1468>.
14. Qin H, Chen F, Huan X, Machida S, Song J, Yuan YA. Structure of the *Arabidopsis thaliana* DCL4 DUF283 domain reveals a noncanonical double-stranded RNA-binding fold for protein-protein interaction. *RNA*. 2010;16(3):474–81. <https://doi.org/10.1261/rna.1965310>.
15. Gholizadeh A. DUF538 protein superfamily is predicted to be chlorophyll hydrolyzing enzymes in plants. *Physiol Mol Biol Plants*. 2016;22(1):77–85. <https://doi.org/10.1007/s12298-015-0331-1>.
16. Ashraf G. Chlorophyll binding ability of non-chloroplastic DUF538 protein superfamily in plants. *Proc Natl Acad Sci India B Biol Sci*. 2018;88(3):967.
17. Zhanguo X, et al. *Arabidopsis* ESK1 encodes a novel regulator of freezing tolerance. *Plant J*. 2007;49(5):786.
18. Kim SJ, Ryu MY, Kim WT. Suppression of *Arabidopsis* RING-DUF1117 E3 ubiquitin ligases, AtRDUF1 and AtRDUF2, reduces tolerance to ABA-mediated drought stress. *Biochem Biophys Res Commun*. 2012;420(1):141–7. <https://doi.org/10.1016/j.bbrc.2012.02.131>.
19. Chiming G, et al. OsSIDP366, a DUF1644 gene, positively regulates responses to drought and salt stresses in rice. *J Integr Plant Biol*. 2016;58(5):492.
20. Min L, et al. Over-expression of a DUF1644 protein gene, SIDP361, enhances tolerance to salt stress in transgenic rice. *J Plant Biol*. 2016;59(1):62.
21. Xiaona H, et al. A novel ABA-responsive TaSRHP gene from wheat contributes to enhanced resistance to salt stress in *Arabidopsis thaliana*. *Plant Mol Biol Report*. 2013;31(4):791.
22. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, et al. The draft genome of a diploid cotton *Gossypium raimondii*. *Nat Genet*. 2012;44(10):1098–103. <https://doi.org/10.1038/ng.2371>.
23. Du X, et al. Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. *Nat Genet*. 2018;50(6):796–802. <https://doi.org/10.1038/s41588-018-0116-x>.
24. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, et al. *Gossypium barbadense* and *Gossypium hirsutum* genomes provide insights into the origin and evolution of allotetraploid cotton. *Nat Genet*. 2019;51(4):739–48. <https://doi.org/10.1038/s41588-019-0371-5>.
25. Wang M, Tu L, Yuan D, Zhu D, Shen C, Li J, et al. Reference genome sequences of two cultivated allotetraploid cottons, *Gossypium hirsutum* and *Gossypium barbadense*. *Nat Genet*. 2019;51(2):224–9. <https://doi.org/10.1038/s41588-018-0282-x>.
26. Zhang M, et al. Risk assessment of cotton cold damage in spring in Xinjiang area. *Jiangsu Agric Sci*. 2020;48(10):102–8.
27. Li H, et al. Occurrence trend of main cotton diseases and insect pests of Xinjiang production and construction corps in 2020. *China Cotton*. 2020;47(04):45–6.
28. Sun F, et al. Comprehensive evaluation of cotton drought tolerance indexes and gray relational analysis. *Agric Res Arid Areas*. 2019;37(01):233–9.
29. Tahir M, et al. Insights into drought stress signaling in plants and the molecular genetic basis of cotton drought tolerance. *Cells*. 2019;9(1):105.
30. Chen Q, et al. Genome-wide identification of cyclophilin gene family in cotton and expression analysis of the fibre development in *Gossypium barbadense*. *Int J Mol Sci*. 2019;20(2):349.
31. Hou C, Tian W, Kleist T, He K, Garcia V, Bai F, et al. DUF221 proteins are a family of osmosensitive calcium-permeable cation channels conserved across eukaryotes. *Cell Res*. 2014;24(5):632–5. <https://doi.org/10.1038/cr.2014.14>.
32. Wing RA, Purugganan MD, Zhang Q. The rice genome revolution: from an ancient grain to green super rice. *Nat Rev Genet*. 2018;19(8):505–17. <https://doi.org/10.1038/s41576-018-0024-z>.
33. Cheng CY, Krishnakumar V, Chan AP, Thibaud-Nissen F, Schobel S, Town CD. Araport11: a complete reannotation of the *Arabidopsis thaliana* reference genome. *Plant J*. 2017;89(4):789–804. <https://doi.org/10.1111/tbj.13415>.
34. Page JT, Liechty ZS, Alexander RH, Clemons K, Hulse-Kemp AM, Ashrafi H, et al. DNA sequence evolution and rare homoeologous conversion in tetraploid cotton. *PLoS Genet*. 2016;12(5):e1006012. <https://doi.org/10.1371/journal.pgen.1006012>.
35. Wang M, Wang P, Lin M, Ye Z, Li G, Tu L, et al. Evolutionary dynamics of 3D genome architecture following polyploidization in cotton. *Nat Plants*. 2018;4(2):90–7. <https://doi.org/10.1038/s41477-017-0096-3>.
36. Li W, Zhang D, Zhu G, Mi X, Guo W. Combining genome-wide and transcriptome-wide analyses reveal the evolutionary conservation and functional diversity of aquaporins in cotton. *BMC Genomics*. 2019;20(1):538. <https://doi.org/10.1186/s12864-019-5928-2>.
37. Waghmare VN, Rong J, Rogers CJ, Bowers JE, Chee PW, Gannaway JR, et al. Comparative transmission genetics of introgressed chromatin in *Gossypium* (cotton) polyploids. *Am J Bot*. 2016;103(4):719–29. <https://doi.org/10.3732/ajb.1500266>.
38. He P, Zhao P, Wang L, Zhang Y, Wang X, Xiao H, et al. The PIN gene family in cotton (*Gossypium hirsutum*): genome-wide identification and gene expression analyses during root development and abiotic stress responses. *BMC Genomics*. 2017;18(1):507. <https://doi.org/10.1186/s12864-017-3901-5>.
39. Shaban M, Ahmed MM, Sun H, Ullah A, Zhu L. Genome-wide identification of lipoxygenase gene family in cotton and functional characterization in response to abiotic stresses. *BMC Genomics*. 2018;19(1):599. <https://doi.org/10.1186/s12864-018-4985-2>.
40. Salih H, Odongo MR, Gong W, He S, du X. Genome-wide analysis of cotton C2H2-zinc finger transcription factor family and their expression analysis during fiber development. *BMC Plant Biol*. 2019;19(1):400. <https://doi.org/10.1186/s12870-019-2003-8>.
41. Roy SW, Gilbert W. The evolution of spliceosomal introns: patterns, puzzles and progress. *Nat Rev Genet*. 2006;7(3):211–21. <https://doi.org/10.1038/nrg1807>.
42. Zhong H, Kong W, Gong Z, Fang X, Deng X, Liu C, et al. Evolutionary analyses reveal diverged patterns of SQUAMOSA promoter binding protein-like (Spl) gene family in *Oryza* genus. *Front Plant Sci*. 2019;10:1–10.
43. Mira MM, Huang S, Hill RD, Stasolla C. Protection of root apex meristem during stress responses. *Plant Signal Behav*. 2018;13(2):e1428517. <https://doi.org/10.1080/15592324.2018.1428517>.
44. Zhang P, Li Q, Zhang C. Root adaptation in *Echinodorus osiris* Rataj plant under cadmium stress. *Int J Phytoremediation*. 2020;22(5):534–9. <https://doi.org/10.1080/15226514.2019.1686605>.
45. Xu J, et al. Proteomic analysis of rapeseed root response to waterlogging stress. *Plants (Basel)*. 2018;7(3):71.
46. Rahamkulov I, Bakhsh A. Tissue-specific and stress-inducible promoters establish their suitability for containment of foreign gene(s) expression in transgenic potatoes. *3 Biotech*. 2020;10(10):426.
47. Neang S, Goto I, Skoulding NS, Cartagena JA, Kano-Nakata M, Yamauchi A, et al. Tissue-specific expression analysis of Na(+) and Cl(−) transporter genes associated with salt removal ability in rice leaf sheath. *BMC Plant Biol*. 2020;20(1):502. <https://doi.org/10.1186/s12870-020-02718-4>.
48. Kuromori T, Seo M, Shinozaki K. ABA transport and plant water stress responses. *Trends Plant Sci*. 2018;23(6):513–22. <https://doi.org/10.1016/j.tplants.2018.04.001>.
49. Raghavendra AS, Gonugunta VK, Christmann A, Grill E. ABA perception and signalling. *Trends Plant Sci*. 2010;15(7):395–401. <https://doi.org/10.1016/j.tplants.2010.04.006>.
50. Tang Y, Zhang Z, Lei Y, Hu G, Liu J, Hao M, et al. Cotton WATs modulate SA biosynthesis and local lignin deposition participating in plant resistance against verticillium dahliae. *Front Plant Sci*. 2019;10:526. <https://doi.org/10.3389/fpls.2019.00526>.
51. Miao Y, Xu L, He X, Zhang L, Shaban M, Zhang X, et al. Suppression of tryptophan synthase activates cotton immunity by triggering cell death via promoting SA synthesis. *Plant J*. 2019;98(2):329–45. <https://doi.org/10.1111/tbj.14222>.
52. Hilaire KT, et al. Impact of elicitation by methyl jasmonate and ethephon on cotton (*Gossypium hirsutum* L.) protection against fusarium oxysporum F. Sp. vasinfectum. *J Adv Agric*. 2018;9:1558–78.
53. Nils S, et al. The PSI family of nuclear proteins is required for growth in *Arabidopsis*. *Plant Mol Biol*. 2014;86(3):289.
54. Yu X, et al. Transcriptome analysis of *Arabidopsis thaliana* and changes of glucosinolates metabolism pathway induced by Flg22. *China Biotechnol*. 2014;34(05):30–8.
55. Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, et al. CottonGen: a genomics, genetics and breeding database for cotton research. *Nucleic Acids Res*. 2014;42(Database issue):D1229–36. <https://doi.org/10.1093/nar/gkt1064>.
56. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. HMMER web server: 2018 update. *Nucleic Acids Res*. 2018;46(W1):W200–4. <https://doi.org/10.1093/nar/gky448>.
57. Lu S, et al. CDD/SPARCLE: the conserved domain database in 2020. *Nucleic Acids Res*. 2020;48(D1):D265.
58. Mariethoz J, et al. Glycomics@EXPasy: bridging the gap. *Mol Cell Proteomics*. 2018;17:2164.

59. Tzu-Hao C, et al. EuLoc: a web-server for accurately predict protein subcellular localization in eukaryotes by incorporating various features of sequence segments into the general form of Chou's PseAAC. *J Comput Aided Mol Des.* 2013;27(1):91.
60. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol.* 2016;33(7):1870.
61. Subramanian B, et al. Evolvview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.* 2019;47(W1):W270–W275.
62. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. *J Hered.* 2002;93(1):77–8.
63. Bailey TL, et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37(Web Server issue):W202–8.
64. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, et al. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13(8):1194–202. <https://doi.org/10.1016/j.molp.2020.06.009>.
65. Lescot M, et al. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* 2002;30(1):325–7. <https://doi.org/10.1093/nar/30.1.325>.
66. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics.* 2018;34(17):i884–90. <https://doi.org/10.1093/bioinformatics/bty560>.
67. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* 2019;37(8):907–15. <https://doi.org/10.1038/s41587-019-0201-4>.
68. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics.* 2014;30(7):923–30. <https://doi.org/10.1093/bioinformatics/btt656>.
69. Tanino Y, Kodama M, Daicho H, Miyauchi Y, Yasumoto T, Yamada Y, et al. Selection of laboratory procedures to detect toxigenic by the 2-step method. *Rinsho Biseibutshu Jinsoku Shindan Kenkyukai Shi.* 2017;27(1):9–14.
70. Otasek D, Morris JH, Bouas J, et al. Cytoscape automation: empowering workflow-based network analysis. *Genome Biol.* 2019;20(1):185.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

