

RESEARCH ARTICLE

Open Access



A comprehensive analysis of cotton VQ gene superfamily reveals their potential and extensive roles in regulating cotton abiotic stress

Pengyun Chen¹, Fei wei^{1,2}, Shuaishuai Cheng^{1,3}, Liang Ma¹, Hantao Wang¹, Meng Zhang¹, Guangzhi Mao¹, Jianhua Lu¹, Pengbo Hao^{1,3}, Adeel Ahmad¹, Lijiao Gu¹, Qiang Ma¹, Aimin Wu¹, Hengling Wei^{1*} and Shuxun Yu^{1*} 

Abstract

Background: Valine-glutamine (VQ) motif-containing proteins play important roles in plant growth, development and abiotic stress response. For many plant species, the VQ genes have been identified and their functions have been described. However, little is known about the origin, evolution, and functions (and underlying mechanisms) of the VQ family genes in cotton.

Results: In this study, we comprehensively analyzed the characteristics of 268 VQ genes from four *Gossypium* genomes and found that the VQ proteins evolved into 10 clades, and each clade had a similar structural and conservative motif. The expansion of the VQ gene was mainly through segmental duplication, followed by dispersal. Expression analysis revealed that many *GhVQs* might play important roles in response to salt and drought stress, and *GhVQ18* and *GhVQ84* were highly expressed under PEG and salt stress. Further analysis showed that *GhVQs* were co-expressed with *GhWRKY* transcription factors (TFs), and microRNAs (miRNAs) could hybridize to their cis-regulatory elements.

Conclusions: The results in this study broaden our understanding of the VQ gene family in plants, and the analysis of the structure, conserved elements, and expression patterns of the VQs provide a solid foundation for exploring their specific functions in cotton responding to abiotic stresses. Our study provides significant insight into the potential functions of VQ genes in cotton.

Keywords: *Gossypium*, Valine glutamine (VQ), Phylogenetic, Expression analysis

Background

The VQ genes form a large gene family with important roles in growth, development and abiotic stress tolerance in plants [1–3]. The VQ proteins have a conserved VQ motif [F**hVQ*hTG (F, phenylalanine; *, any amino acid; h, hydrophobic residue; V, valine; Q, glutamine; T,

tryptophan; G, glycine)] [4, 5] and interact with WRKY TFs via the conserved residues V and Q. In *Arabidopsis thaliana*, many VQ genes have been reported to function in development and responses to abiotic stress. For example, *AtVQ23* (sigma factor-binding protein 1, *SIB1*), and *AtVQ16* (*SIB2*) were found to interact with *AtWRKY33* to increase the resistance of *Arabidopsis* plants to *Botrytis cinerea* [6]. In another study, *AtVQ16* and *AtVQ23* have also been proven could interact with *AtWRKY57*, and *AtVQ16* and *AtVQ23* can enhance the

* Correspondence: henglingwei@163.com; ysx195311@163.com

¹State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China
Full list of author information is available at the end of the article



© The Author(s). 2020 **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.

competitions on *AtWRKY57* to *AtWRKY33* in regulating *JASMONATE ZIM-DOMAIN1* (*JAZ1*) and *JAZ5* [7]. Moreover, *JASMONATE-ASSOCIATED VQ MOTIF GENE1* (*JAV1/AtVQ22*) has been addressed to be as a key negative regulator of the jasmonate signalling [8]. For instance, *AtVQ09* acts as a repressor of *AtWRKY08* factor to modulate salt tolerance in *Arabidopsis* [9]. Recently, *MdVQ10* and *MdVQ15* were also described to interact with *MdWRKY52* to regulate pathogen defense and development in apple (*Malus domestica Borkh*) [10]. In addition, *OsVQ7* interacts with *OsWRKY24* and play roles in NO signaling contributing to the tolerance of various stresses and development in rice (*Oryza sativa*) [11].

VQs have also been reported to perform other functions. It has been demonstrated that *AtVQ14* (*HAIKU, IKU1*) interacts with *MINISEED3* (*MINI3, AtWRKY10*) to reduce the expression of *IKU2*, affecting the seed size [1], and *AtVQ20* regulates the male gametogenesis in *Arabidopsis* [12]. In addition, some VQs have been shown to interact with ETHYLENE RESPONSE FACTORS (*ERF*), mitogen-activated protein kinases (*MAPKs*), and miRNAs in response to environmental stresses in plants [13, 14]. Otherwise, some abiotic stress-related genes have been isolated from *Arabidopsis* and other plants [15–19], including VQ TFs, *WRKY* TFs, and other TFs.

Cotton is an important, widely cultivated fiber and oil crop that is essential for the textile industry and provides nutrient-rich edible oil [20]. Various biotic and abiotic stresses, including pathogen infection, drought and salinity stresses, consistently and severely affect the formation of cotton production [21–23]. Therefore, it is very important to develop new cultivars with high resistance to biotic and abiotic stresses. Recently, VQ family genes have been identified at genome-wide levels in several plants, including *Arabidopsis* (34 VQs) [24], soybean (*Glycine max*) (74 VQs) [25], rice (39 VQs) [26], maize (*Zea mays*) (61 VQs) [27], Chinese cabbage (*Brassica rapa* spp. *pekinensis*) (57 VQs) [28], apple (49 VQs) [10], tea (*Camellia sinensis*) (25 VQs) [29], tomato (*Solanum lycopersicum*) (26 VQs) [30], sunflower (*Helianthus annuus*) (20 VQs) [31], tobacco (*Nicotiana tabacum*) (59 VQs) [19], *Cicer arietinum* (19 VQs) [32] and *Medicago truncatula* (32 VQs) [32]. However, the genomic information and genetic evolution relationships of VQs are not clear in *Gossypium* spp., and the expression patterns of these genes in different tissues and in responses to abiotic stresses remain unknown. The released for cultivated *Gossypium* spp. (*G. hirsutum* Linn., *G. barbadense* L., *G. raimondii* Ulbr and *G. arboretum* L.) genome sequences and their annotation have brought much convenience for thoroughly investigation of their genomics and genetic information [33–35]. In this study, using the

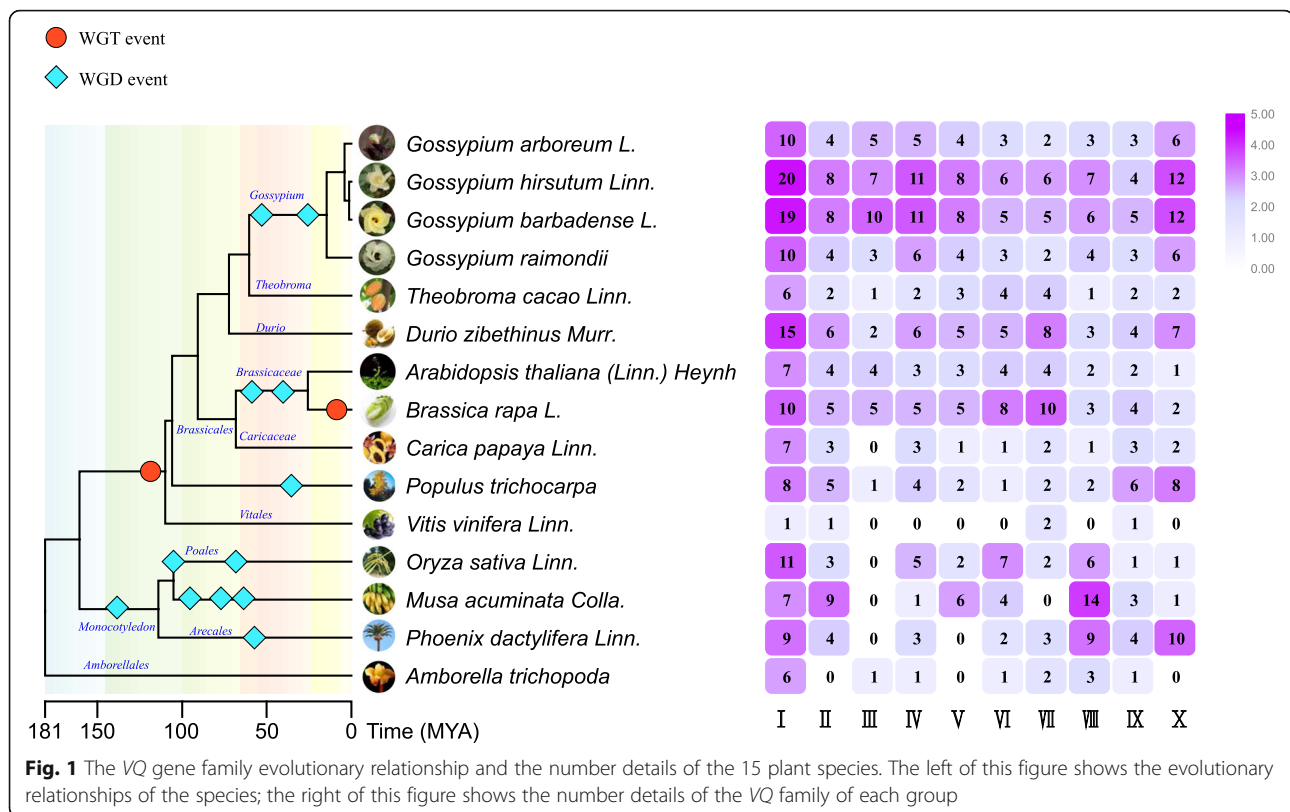
annotations of four *Gossypium* genomes, we identified the VQs of these *Gossypium* species, performed the analyses of phylogeny, conserved structural motifs, whole-genome duplication (WGD) and functional interaction networks of the VQs, and predicted the microRNA target profiles of the VQs. The comprehensive analysis of the VQ gene family in *Gossypium* will contribute to identify new key candidate genes for diverse stress resistance in cotton breeding.

Results

Identification and comparative analysis of VQs in plants

To identify VQ family in *G. hirsutum*, *G. barbadense*, *G. raimondii* and *G. arboretum*, the AtVQ proteins were used as query sequences to search against the protein databases of the four *Gossypium* species, and the VQ-domain Pfam (PF05678) was also applied. In total, 89 *GhVQs*, 89 *GbVQs*, 45 *GrVQs*, and 45 *GaVQs* were identified and named in *G. hirsutum*, *G. barbadense*, *G. raimondii* and *G. arboretum*, respectively (Additional File 1, Supplemental Table S1). In addition, the physiological and biochemical properties of 268 VQs in *Gossypium* species were determined, including CDS length, GC count, isoelectric point (pI) and molecular weight (MW) (Additional File 1, Supplemental Table S1). The CDS lengths of these *Gossypium* VQs ranged from 279 bp (*GhVQ89* and *GbVQ89*) to 1443 bp (*GbVQ15*), the average GC content of the transcripts was 46.01, their exon numbers varied from 1 to 9, and only a small percentage of VQs contained introns (3.37% *GhVQs*, 3.37% *GbVQs*, 6.67% *GaVQs*, and 31.11% *GrVQs*). The pI values varied from 4.159 (*GbVQ33* and *GbVQ78*) to 11.496 (*GhVQ07*) and the MW values ranged from 10.346 kDa (*GbVQ89*) to 52.058 kDa (*GbVQ15*) (Additional File 2, Supplemental Fig. S1).

To perform comparative genomic analyses, we searched another 11 plant species for VQ proteins. The evolutionary relationships of the species and the number of their VQ genes are shown in Fig. 1. The data revealed that the numbers of VQs in *A. trichopoda*, *P. dactylifera*, *V. vinifera*, *P. trichocarpa*, and *T. cacao* were less than those in the four *Gossypium* species (Additional File 3, Supplemental Table S2). The comparative structure analysis of VQs showed that almost all the VQs had a few introns and encoded relatively small proteins, and only 3 *GhVQs*, 3 *GbVQs*, 3 *GaVQs* and 14 *GrVQs* had more than one intron. We speculated that the WGD events that occurred during the evolution of angiosperms increased the numbers of the *Gossypium* VQs, and these events helped the VQs to gain new functions through neofunctionalization. However, the evolutionary forces that shaped the current intron/exon gene structures remain unknown.



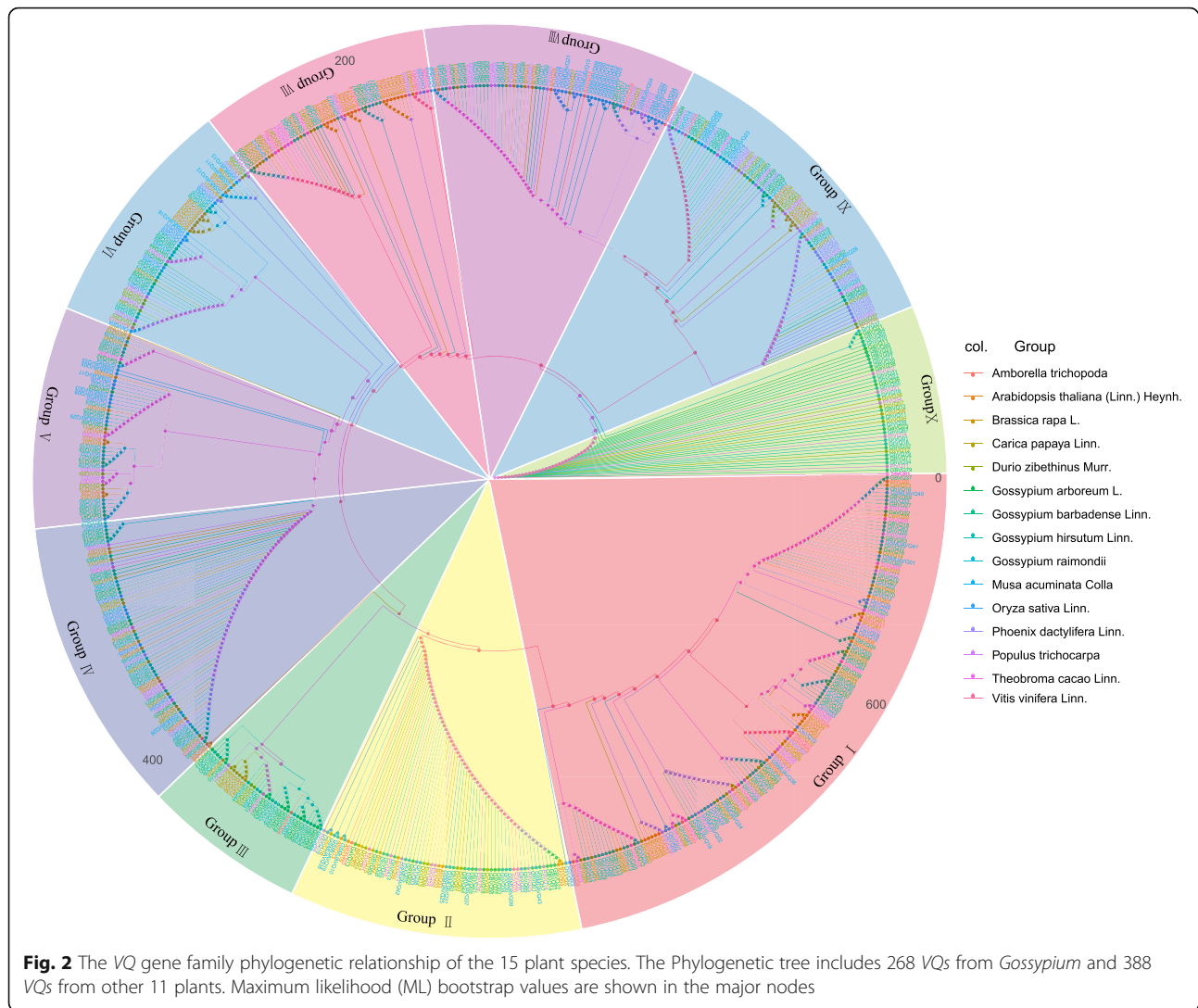
Phylogenetic analysis of VQs

To explore the relationships among VQs in *Gossypium*, we conducted a phylogenetic analysis of the VQs from the 15 plant species (Fig. 2), and a phylogenetic tree between *Gossypium* spp. and *Arabidopsis* was also constructed (Additional File 4, Supplemental Fig. S2). The tree contained 656 VQs. These proteins were divided into 10 clades based on the nomenclature of the *Arabidopsis* VQs. The largest group (Group I) contained 20 GhVQs, 20 GbVQs, 10 GaVQs, and 10 GrVQs. Group IX was the smallest group, including 4 GhVQs, 5 GbVQs, 3 GaVQs and 3 GrVQs. Previous research has verified that VQ proteins contain a conserved motif composed of F**hVQ*hTG [4, 5]. In our study, among the 656 VQs, 212 proteins (in Group I and Group II) had the amino acid “M” next to “VQ” (simple M-VQ model); 159 proteins (in Group III, Group IV, and Group VII) had the amino acid “V” next to “VQ” (simple V-VQ model); and 285 proteins (in Group V, Group VI, Group VII, Group IX and Group X) had the amino acid “L” next to “VQ” (simple L-VQ model) (Additional File 5, Supplemental Fig. S3). The VQs with rarer amino acids of the *Gossypium* species were also scattered in Group I to Group X, and the clusters of VQs were similar to those in angiosperms [3].

Cis-regulation elements and structural composition of the VQs

The cis-regulation elements in the promoters (from 2000 bp to -1 bp) of the four *Gossypium* VQs were analyzed using the PlantCARE tool. 715, 701, 386 and 399 cis-regulation elements from the GhVQs, GbVQs, GaVQs and GrVQs were identified, respectively. Among these, seven kinds of hormone-responsive cis-regulation elements, ABRE (ABA-responsive element), P-box, TGA-box, TGA-element, TCA-element, CGTCA-motif, and GARE-motif were associated with ABA, ethylene, salicylic acid (SA), methyl jasmonate (MeJA), auxin (IAA), and gibberellin (GA), respectively; and six regulatory elements, MBS, TC-rich repeats, LTR, DRE-motif, W-box, and CCAAT-box were related to drought, cold stress, and pathogen defense (Fig. 3). Moreover, the promoters of 66 GhVQs, 69 GbVQs, 36 GrVQs and 34 GaVQs possessed WRKY-binding sites (W-box) (Fig. 3). The diversity of the cis-regulation elements in the promoters of VQs indicated that VQs might participate in regulating the *Gossypium* response to endogenous hormones and diverse environmental stimuli.

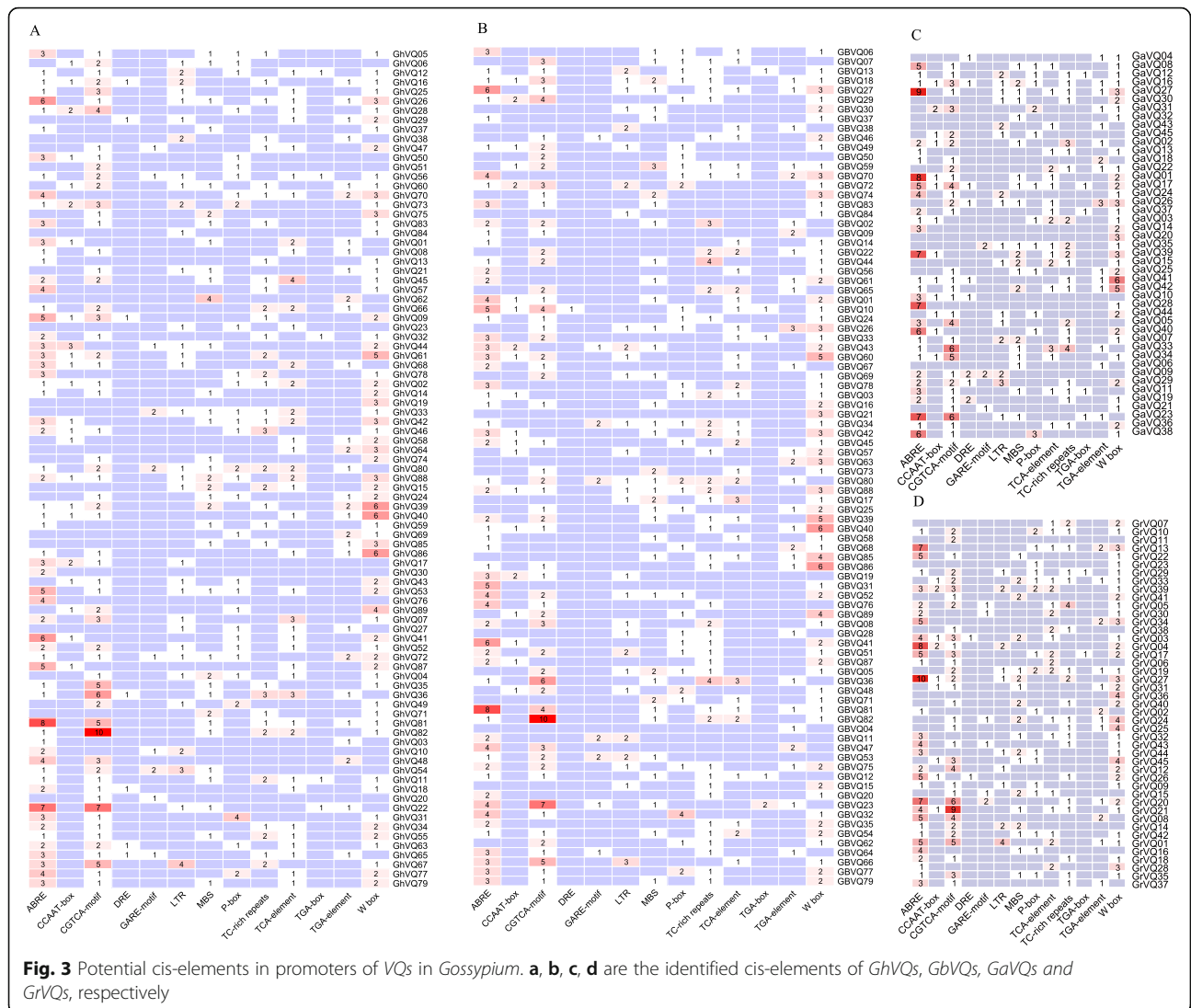
Motif compositions and exon-intron structures of the VQs were shown in Fig. 4. Combining the phylogenetic data groups of the four *Gossypium* VQs, we



found that there were more motif types in Group IV, including Motif 1, Motif 4, Motif 7, Motif 9, and Motif 10; and followed by Group II, containing Motif 1, Motif 2, Motif 3, Motif 6, and Motif 8. Not surprisingly, Motif 1 existed in almost all of the VQs, suggesting that it was the most conservative motif. The differences in motif composition among the four *Gossypium* VQs suggested that they might perform different functions in diverse *Gossypium* species. Most of the VQs had no intron, including 96.63% (86/89) of *Gh*VQs, 96.63% (86/89) of *Gb*VQs, 68.89% (31/45) of *Gr*VQs, and 93.33% (42/45) of *Ga*VQs. The remaining VQs, which were widely distributed in Group VII and Group X, contained one to eight introns. In general, VQs in the same clades would share similar motif elements and structural compositions, indicating that the VQ members in the same subgroup could have similar functions.

Chromosomal distribution, synteny and duplication analysis

In this study, VQs were detected located in most chromosomes with a few exceptions. For example, Gh_A09 and Gh_D02 were in *G. hirsutum*, Gb_A09, Gb_A13 and Gb_D02 were in *G. barbadense*, Ga_Ch03 and Ga_Ch09 in *G. arboretum*, and Gr_Ch05 were *G. raimondii* (Fig. 5). For the two allotetraploid species of *Gossypium*, Gh_D05 (eight genes / ~ 9%), Gb_A05 (eight genes / ~ 9%), and Gb_D05 (eight genes / ~ 9%) contained more VQ genes than other chromosomes, while Gh_A02, Gh_A03, Gh_A13, Gh_D08, Gh_D13, Gb_A02, Gb_A03, Gb_D08, Gb_D09, and Gb_D13 only contained one gene. For the two diploid species, Ga_Ch05 (eight genes / ~ 17.8%), Gr_Ch02 (seven genes / ~ 17.8%), Gr_Ch07 (seven genes / ~ 17.8%) and Gr_Ch09 (seven genes / ~ 17.8%) contained more VQs, and Ga_Ch08, Ga_Ch13, Gr_Ch04, Gr_Ch12, and Gr_Ch13 only

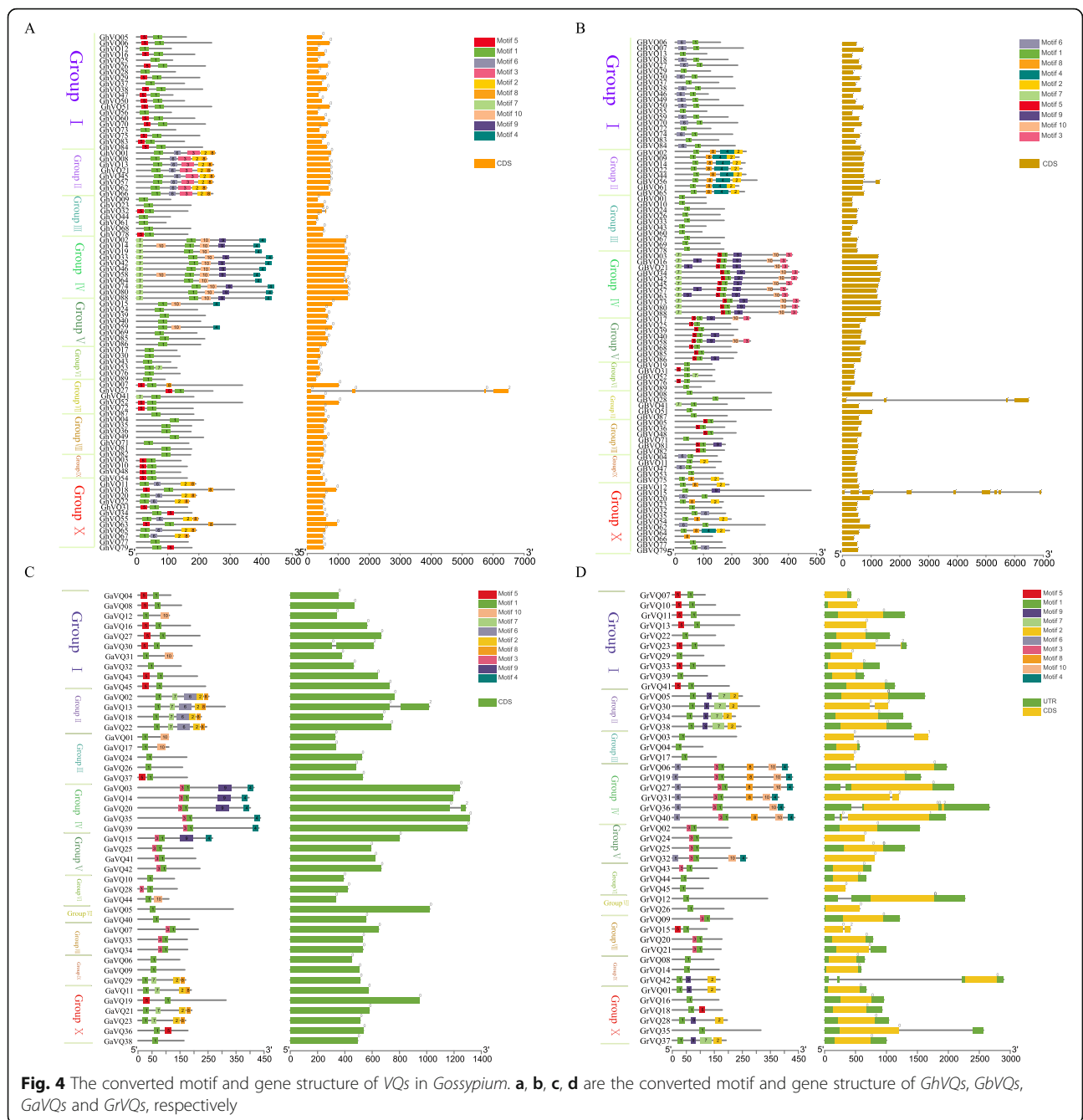


contained one gene. Most VQs in the four *Gossypium* species were distributed at both ends of the chromosomes, which corresponded to the position of the telomere.

The orthologous VQs were first identified between *G. hirsutum* and *G. arboreum* with *G. raimondii*. A total of 83 GhVQs were orthologous genes in the two-diploid *Gossypium*, of which 40 gene pairs showed A genome origin, while 43 gene pairs showed D genome origin (Additional Files 6 and 7, Supplemental Fig. S4 and S7, Table S3). Subsequently, orthologous gene identification was also conducted between *G. barbadense* and *G. arboreum* with *G. raimondii*, and there were 84 orthologous GbVQs. Of which, 40 gene pairs showed A genome origin, while 44 gene pairs showed D genome origin (Additional Files 7 and 8, Supplemental Fig. S5 and S3). Orthologous genes between *G. hirsutum* and *G. barbadense* were also identified. It was found that 39 gene

pairs in Gh_At and Gb_At subgenomes and 42 gene pairs in Gh_Dt and Gb_Dt subgenomes (Fig. 6 and Additional File 7, Supplemental Table S3). In addition, GhVQ27, GhVQ61, GhVQ68, GbVQ28, GbVQ60 and GbVQ67 had no orthologous genes in the diploid *Gossypium* species.

As previously described, duplication contributed to the expansion of genes in the polyploid events in plants [36]. The tetraploid species *Gossypium* of have undergone a genome duplication since their diverge from the two diploid species of *Gossypium*. In our study, we have identified the VQ duplication event, and the WGD/segmental event likely contributed to the expression regulation of VQs in *Gossypium*. The percentages of VQs derived from WGD were 60.47% in the At-subgenome of *G. hirsutum*, 63.04% in the Dt-subgenome of *G. hirsutum*, 69.04% in the At-subgenome of *G. barbadense*, 59.57% in the Dt-subgenome of *G. barbadense*, 62.22%

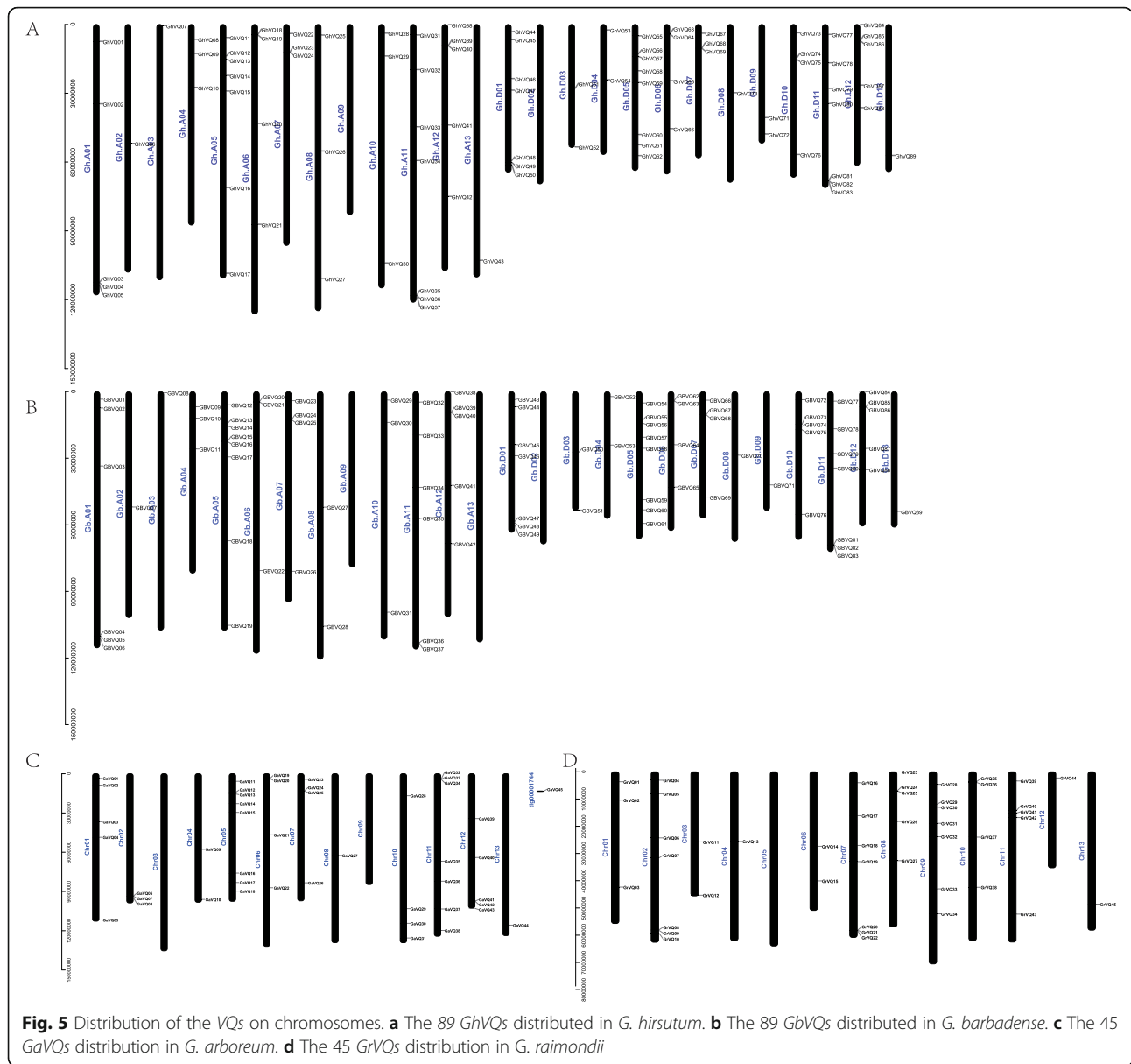


in *G. raimondii*, and 57.78% in *G. arboreum* (Additional File 9, Supplemental Table S4). Gene duplication events after the divergence of *Gossypium* species resulted in a high number of paralogous genes in both allotetraploid *Gossypium* species.

Prediction of miRNA target sites

miRNA had been predicted to target the VQ genes in *Arabidopsis* [37, 38] and tea [39]. To determine the miRNA-mediated post-transcriptional regulation of VQs

in two allotetraploid species of *Gossypium*, we predicted the target sites of miRNAs in the coding (CDS) regions of the *GhVQs* and *GbVQs*. In *G. hirsutum*, 46 sites of 34 *GhVQs* were detected that could be targeted by 22 miRNAs, while 46 sites of 32 *GbVQs* could be targeted by 21 miRNAs (Fig. 6 and Additional File 10, Supplemental Table S5). Of these, six VQ genes (*GhVQ02*, *GhVQ40*, *GhVQ86*, *GbVQ02*, *GbVQ40* and *GbVQ86*) were predicted to be targeted by Ghr-miR172 in the CDS regions; and six VQ genes (*GhVQ39*, *GhVQ52*, *GhVQ85*,

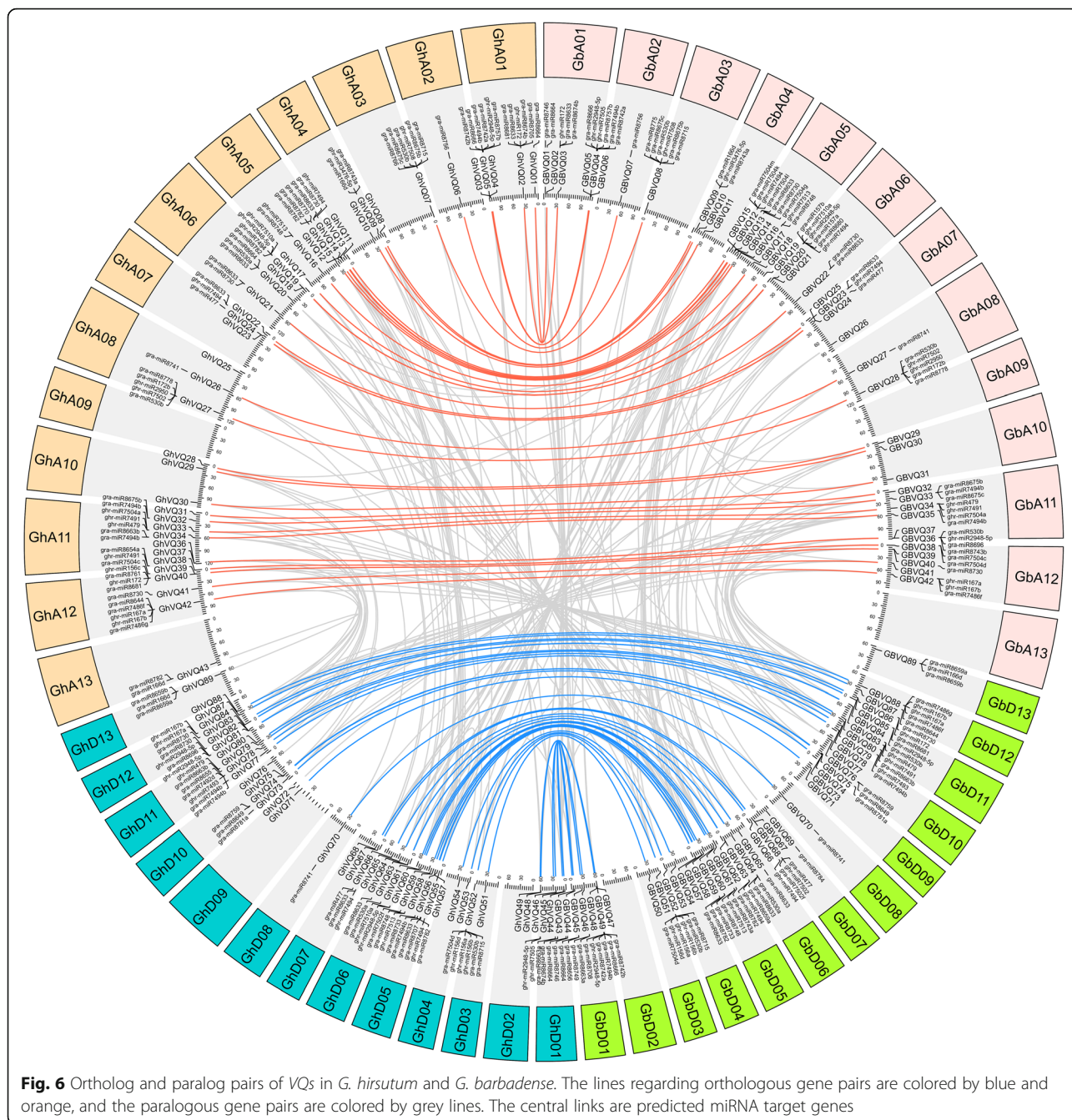


GbVQ39, *GbVQ51* and *GbVQ85*) were targeted by Ghr-miR156 (Ghr-miR156a, Ghr-miR156b, Ghr-miR156c and Ghr-miR156d) at 10 prediction sites. Ghr-miR172 and Ghr-miR156 were reported to be involved in some biological processes, including the responses to developmental cues and abiotic stress in plants [40–42]. However, it requires further experiments to verify the regulation mechanism and functions of those predicted miRNAs and their targets in *Gossypium*.

Expression pattern analysis and function verification

Expression profiles of the VQs in the two allotetraploid kinds of *Gossypium* were analyzed with available transcriptome data (Additional Files 11 and 12,

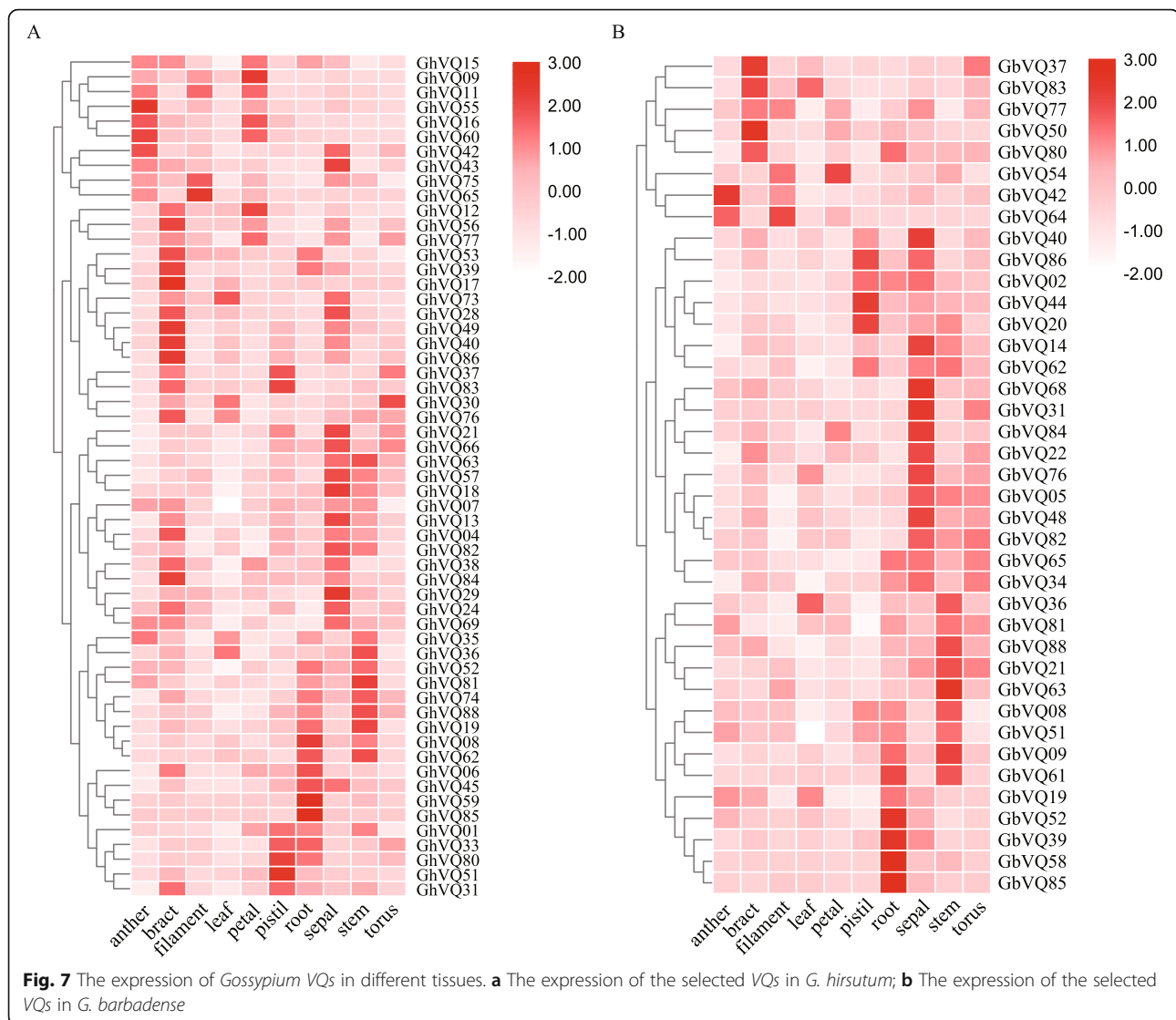
Supplemental Fig. S6 and S7 and Supplemental Table S6). In this study, the *GhVQs* and *GbVQs* with average FPKM values > 5 and being present in at least two samples were identified as potentially expressed transcripts. Fifty-seven *GhVQs* and 39 *GbVQs* were selected, and their expression profiles tested in 10 tissues, including the anther, bract, filament, leaf, petal, pistil, root, sepal, stem and torus (Fig. 7). For 57 *GhVQs*, 22 genes were highly expressed in roots and *GhVQ82* had the highest expression level, and 14 genes were highly expressed in leaves. There were only a few genes expressed in the anther, bract, filament, petal, pistil, sepal, stem and torus (Fig. 7a). For 39 *GbVQs*, 17 *GbVQs* were strongly highly expressed in roots, and 12 genes were strongly expressed



in leaves (Fig. 7b). The different expression profiles of VQ genes suggest that they have different functions in distinct tissues and developmental stages.

VQs were widely identified in the abiotic stress responses in angiosperms [3, 32]. In this study, the expression patterns of VQs in the allotetraploid *Gossypium* types under salt, drought, cold and heat stresses were analyzed using the published data. In total, 43 *GhVQs* and 37 *GbVQs* had different expression profiles under the four abiotic stress treatments (Fig. 8). Under salt stress, 29 *GhVQs* were significantly up-regulated at 12 h,

and 21 *GbVQs* were up-regulated at 6 h (Fig. 8a and c). Upon PEG treatment, most of the *GhVQs* and *GbVQs* were highly expressed at 12 h (Fig. 8b and d). During cold stress, 33 *GhVQs* and 19 *GbVQs* were up-regulated at 24 h (Fig. 8e and g). Most of the *GhVQs* and *GbVQs* under the hot treatment were highly expressed at 1 h (Fig. 8f and h). To validate of the expression results of *GhVQs* in response to salt and drought stresses, we conducted qRT-PCR analyses of 12 *GhVQs* after treatments with PEG and salt treatment. In the presence of PEG, *GhVQ08*, *GhVQ18*, *GhVQ62*, *GhVQ64*, *GhVQ80*, and



GhVQ84 had high expression levels at 48 h, while these *GhVQs* except *GhVQ18* and *GhVQ84*, were highly expressed during 24–48 h under salt treatment (Fig. 9). The qRT-PCR results were slightly different from the RNA-seq data, but these findings suggest that some *GhVQs* were involved in plant response to drought and salt stresses.

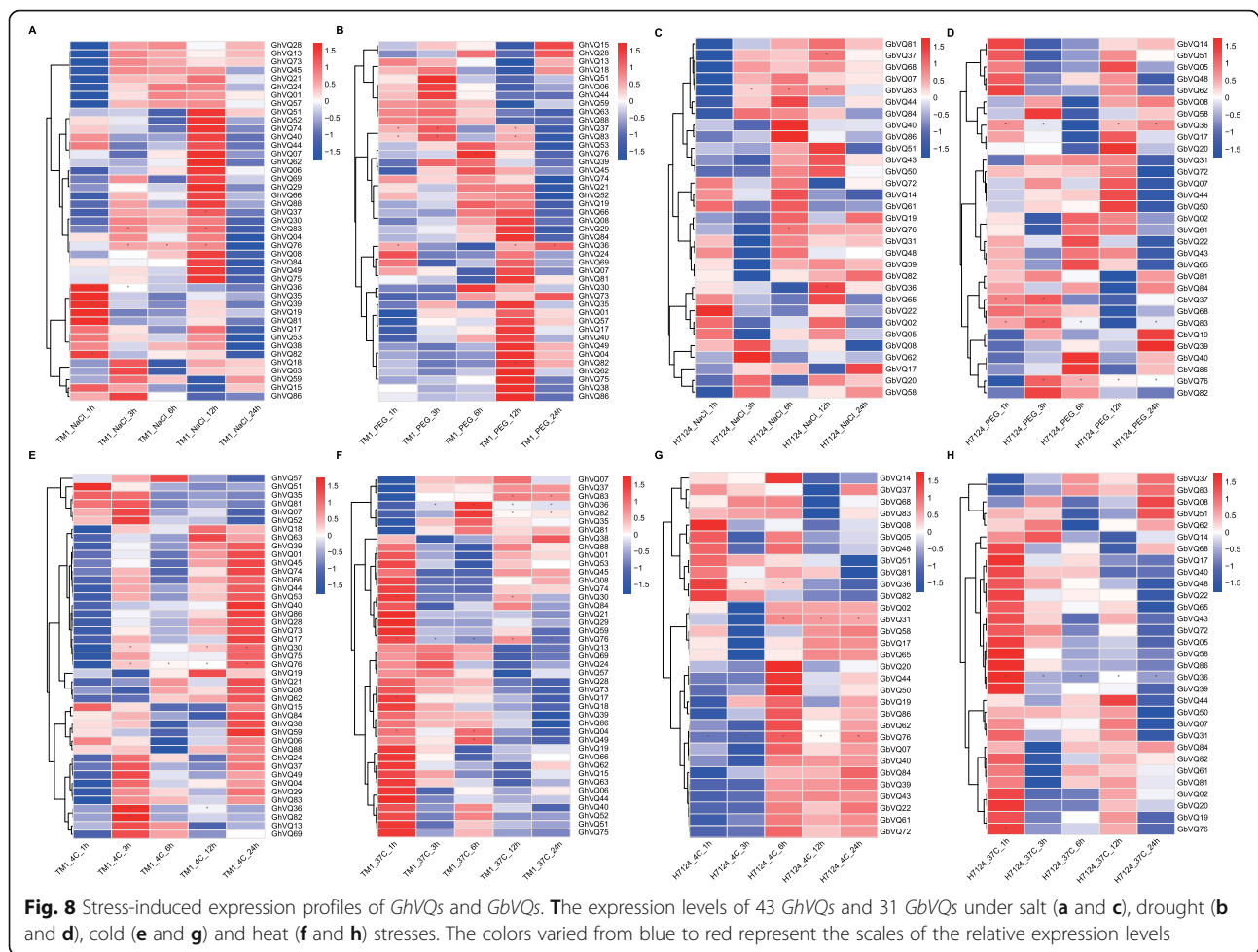
Co-expression and interaction networks of *GhVQs*

To understand the putative roles of VQs in plant adaptation to drought and salt stresses, we conducted a co-expression analysis. Ten *GhVQs* were found to co-express with another 227 functional genes (Fig. 10 and Additional File 13, Supplemental Table S7). Among these, six and seven VQs were identified in different modules of drought stress and salt stress, respectively, while *GhVQ37*, *GhVQ59* and *GhVQ83* were detected coexisting during the two stress treatments. Moreover,

these 227 genes co-expressing with 10 *GhVQs*, contained multiple TFs, including domain AP2, bHLH, F-box, GRAS, p450, PLB03212, WD40 and WRKY (Fig. 10a–e and Additional File 13, Supplemental Table S7). The functional regulation networks of the *GhVQs* were constructed using the website of STRING11.0 with the module reference of *Arabidopsis* association, and the results revealed that the *GhVQs* participated in plant defense interaction networks, including WRKYs, MYB15, MPK4, AR781, CSN5B and SIGAs (Fig. 10f). Indeed, VQ proteins could interact with WRKYs and other TFs to defend against abiotic stresses in cotton.

Discussion

In previous studies, the VQ family genes have been systematically analyzed in *Arabidopsis* [24], soybean [25], sunflower [31], rice [26], banana [43], maize [27], bamboo [44], *Cicer arietinum* [32], *Medicago truncatula*



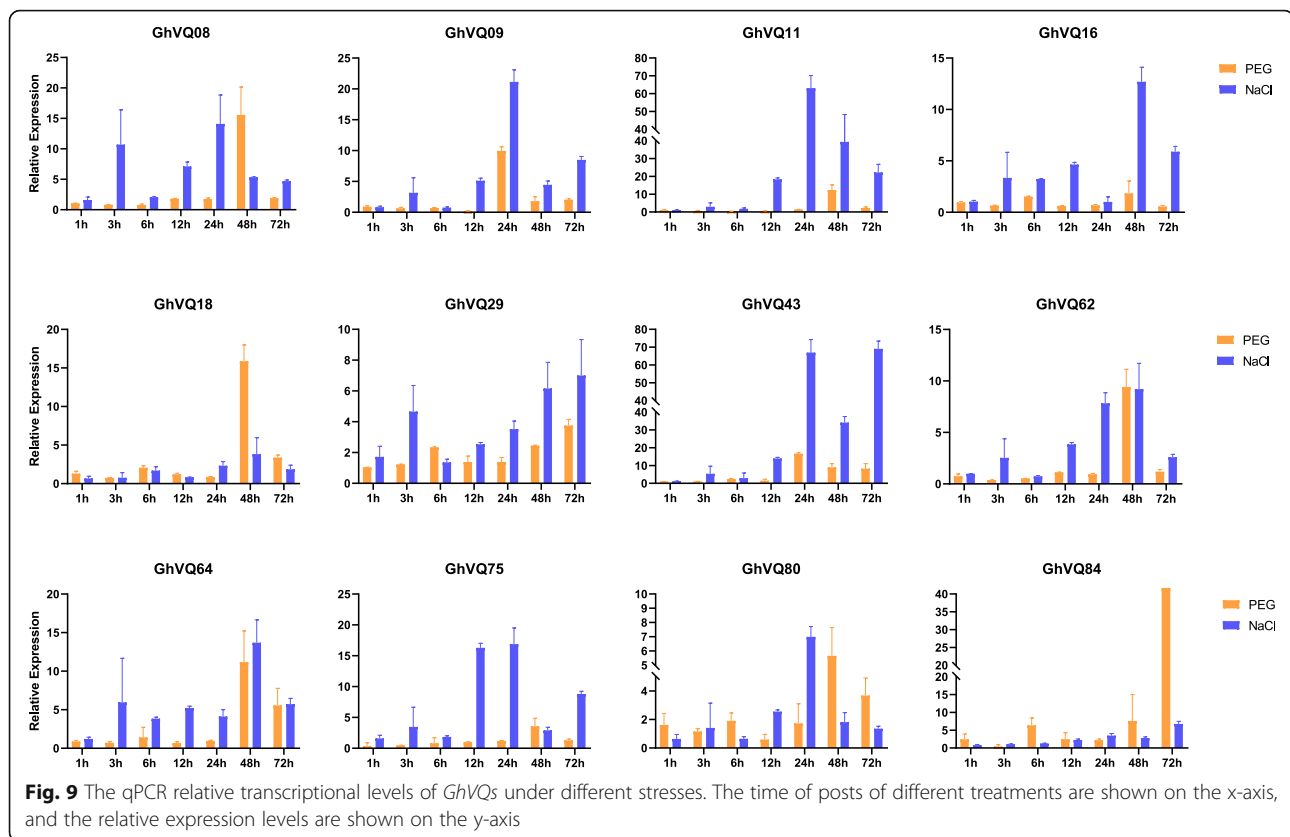
[32], and tobacco [19], and has shown to play significant roles in regulating growth, development processes, and responding to biotic and abiotic stresses [2]. Here, we complete a comprehensive analysis of the VQ family genes and explore their evolutionary mechanism in *Gossypium* species.

The expansion, duplication and structural characteristics of VQs in *Gossypium*

In this study, we analyzed the VQs of *G. hirsutum*, *G. barbadense*, *G. raimondii*, *G. arboretum* and another 11 plant species, and found that the number of VQs in the genomes of 15 species was inconsistent with related to the size of their genomes. There are 89 *GhVQs*, 89 *GbVQs*, 45 *GaVQs*, and 45 *GrVQs*, respectively. The number of *Gossypium* VQs was higher than that in *cacao* (27 *TcVQs*) and in *Arabidopsis* (34 *AtVQs*), but the number of *GrVQs* and *GaVQs* was fewer than the *DzVQs* (Fig. 1 and Additional File 2, Supplemental Table S2). Previous studies have shown that diverse WGD events led to the different sizes of plant genomes [45–

47]. Our results indicated that VQs in these four *Gossypium* species were more likely to be proximal, tandem, and segmental genes, while the majority of VQs in rice [26] and *Arabidopsis* [24] are singleton genes. Through the analysis of the phylogenetic and structural features of the 15 plants VQ domains, the VQs could be divided into 10 clades. Group III could be expanded in the eudicots, particularly in the Mallow species (Fig. 2), while Groups III, IV, V, VI, VII, and X had no *VvVQs*, suggesting that these might have been lost in ancient genome duplication events.

Ten conserved motifs were also identified in the four *Gossypium* VQs. Motif 1 corresponded with the VQ-containing motif, which is widely found in angiosperms [3]. Previous studies have suggested that VQs have few introns in higher plants, being agreement with the results from *Gossypium* VQs. Only 3 *GhVQs*, 3 *GbVQs*, 14 *GrVQs*, and 3 *GaVQs* had multiple introns. Additionally, the motif compositions and intron contents of the *Gossypium* VQ proteins/genes in our study were consistent with the results of the phylogenetic analysis and the type



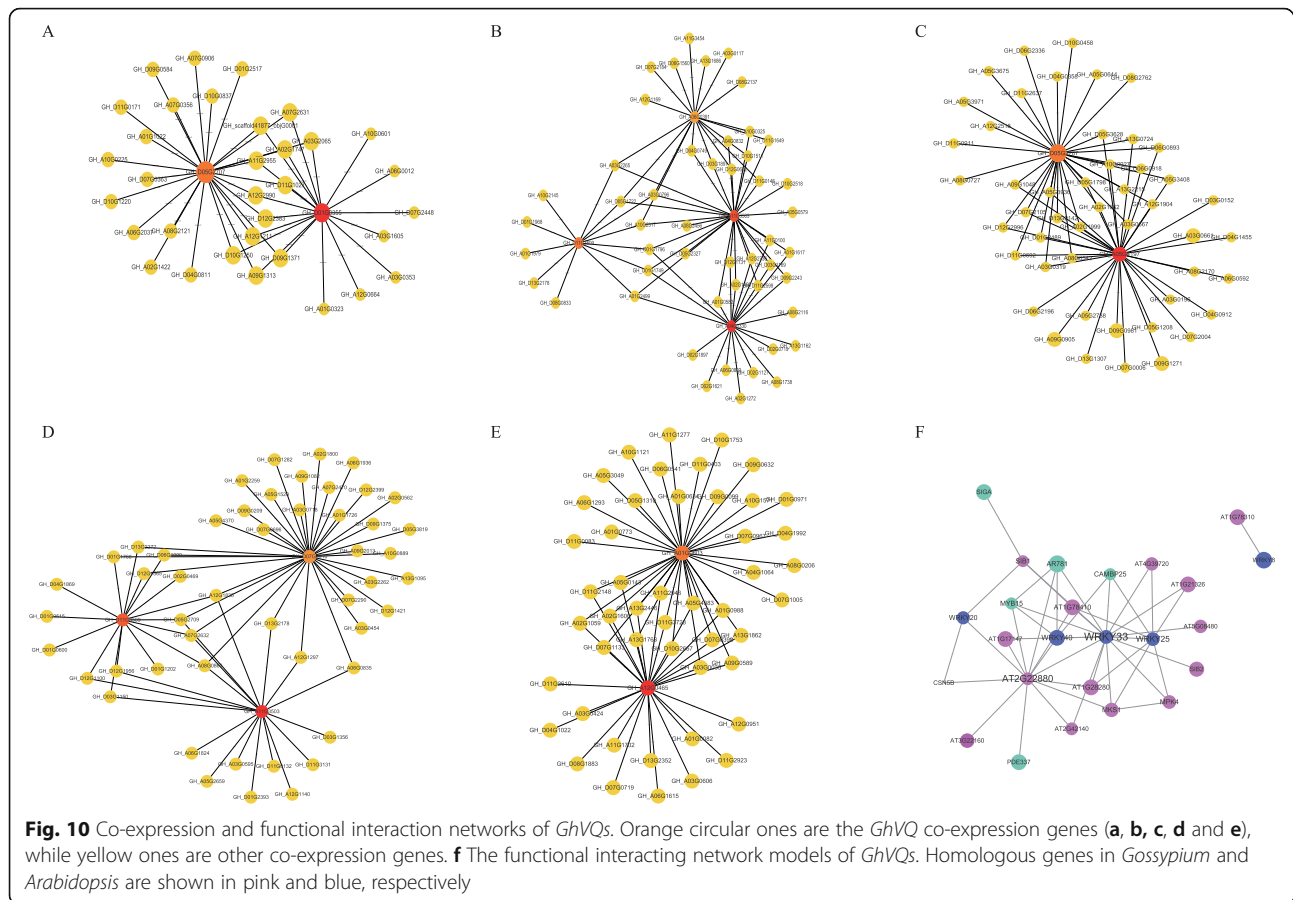
of gene duplication. Collectively, our data speculated that *Gossypium VQs* might be affected by intronic evolution.

VQs play important roles in abiotic stress signaling pathways

Previous reports have shown that VQs are involved in various endogenous and environmental signals, which consistent with their diverse roles in plant development and in the response to abiotic stresses [2, 5, 19, 24, 26, 32]. For example, *AtVQ08*, *AtVQ14*, *AtVQ17*, *AtVQ18* and *AtVQ22* are involved in modulation of seed development, chloroplast development, and plant growth. A proportion of the *GmVQs* [25], *PeVQs* [44], *VvVQs* [48], *CsVQs* [29], *HaVQs* [31], and *NtVQs* [19] also function in regulating the growth of different tissues at different developmental stages. In this study, most *GhVQs* and *GbVQs* were found to differentially expressed in the different tissues, including the ovule, fiber, anther, leaf, root, sepal and stem, suggesting that the VQs may play an important role in growth and development of *Gossypium* species (Fig. 7, Additional Files 6 and 7, Supplemental Fig. S6 and S7). Most VQs have been demonstrated to play important roles in responses to various abiotic stresses in plants [2, 3, 19, 27, 30, 32]. In

our work, we assessed the expression levels of the *GhVQs* and *GbVQs* under salt, drought, cold, and heat stresses, and found that the majority of the VQs were up-regulated under drought, salt, and cold stress, or down-regulated under heat stress. These findings were similar to those of previous reports in *Arabidopsis* [5], rice [26], maize [27] and cabbage [28]. Also, the promoters of the *Gossypium VQs*, many cis-elements that were reported to exist in other abiotic stresses responsive genes were detected (Fig. 3), implying that the VQs in *Gossypium* are likely involved in response to various abiotic stresses, and that the response mechanisms maybe complex and diverse.

VQs has been reported to interact with WRKY TFs and to regulate a variety of physiological and biochemical processes and abiotic stress responses [6, 9, 11, 49, 50]. Here, by constructing co-expression and an *Arabidopsis* associated model, multiple *Gossypium VQs*, such as *GhVQ37*, *GhVQ59*, and *GhVQ83*, were predicted to interact with different WRKY TFs, implying that VQs act in stress tolerance through interacting with WRKYs. Moreover, we predicted some putative target sites of microRNAs in the *Gossypium VQs*. These microRNAs included miRNA156s and miR172, which serve important roles in various life processes of plants [40–42].



These results indicate that the *Gossypium* VQs are extensively involved in growth, development, and in response to stresses, and work together with WRKYs and microRNAs during these processes.

Conclusions

In this study, using bioinformatics plus expression profiles, we identified and presented the structure, phylogenetic relationships, and tissue specificity of VQ family genes in four *Gossypium* species. Our data showed that the gene structure and motif coding regions were conserved across plants, and segmental, dispersed, and tandem duplications were the main reason for the expansion of the VQs. Cis-element and expression analyses indicated that the majority of VQs were activated in response to abiotic stress, and some of VQs were co-expressed with WRKYs and hybridized with the miRNAs involved in *Gossypium* growth, development, and abiotic stress. Our study could serve as a foundation for future exploration of the specific functions of *Gossypium* VQs in the abiotic stress responses and the interactions with WRKYs or microRNAs.

Methods

Identification and classification of VQs in plants

The latest versions of predicted proteomes of *G. raimondii* (https://phytozome.jgi.doe.gov/pz/portal.html#info?alias=Org_Graimondii) [35], *G. arboreum* (<ftp://bioinfo.ayit.edu.cn/downloads/>) [34], *G. hirsutum*, and *G. barbadense* (<https://ibi.zju.edu.cn/cotton/>) [33] were used in this study. The genome data of other plants were obtained from the JGI database (<https://www.phytozome.net>) and National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/>): *A. thaliana*, *V. vinifera*, *T. cacao*, *P. trichocarpa*, *P. dactylifera*, *O. sativa*, *M. acuminata*, *D. zibethinus*, *C. papaya*, *B. rapacious*, and *A. trichopoda*. The pre-classified groups of these species were based on their phylogenetic relationships (<https://www.timetree.org/>) [51]. The VQ conserved domain (PF05678) was used as a query to scan the *Gossypium* species protein databases, and the *A. thaliana* VQ proteins were used as the queries to search against the above proteomes through the basic local alignment search tool (BLAST, v 2.10.0) (score value ≥ 0.0001 and E-value = 1×10^{-3}) [52] for each newly identified gene. The obtained putative VQ

motif-containing sequences were confirmed in the NCBI Conserved Domain Database (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>) [53] and SMART database (<https://smart.embl-heidelberg.de/>) [54]. Then, the physical and chemical properties of the VQ family members, including amino acid length, mRNA length, MW, and pI, were analyzed using the online tools of the web site of softberry website (<https://linux1.softberry.com/berry.phtml>) and the ExPASy website (<https://web.expasy.org/translate/>) [55], and the relative results were plotted by the ggstatsplot (v 0.4.0) [56].

Phylogenetic and synteny analysis of the VQs in plants

All the VQ motif-containing proteins from the four *Gossypium* and other 11 plant species were aligned using MAFFT (Multiple Alignment using Fast Fourier Transform, v 7.4.0.7) (L-INS-algorithms) [57] with default parameters, and conserved site sequences were selected by the Gblock (v 0.91b) software (https://molevol.cmima.csic.es/castresana/Gblocks_server.html) [58]. A phylogenetic tree was constructed using the IQ-TREE software (v 1.6.9) (<https://www.iqtree.org/>) [59] with the maximum likelihood method, and the substitution model was calculated with ModelFinder (intergraded in IQ-TREE; best-fit model: JTT + R5 chosen according to BIC). The obtained treefile was visualized using the ggtree (v 2.0.2) [60] and AI (Adobe Illustrator CS6).

The synteny and collinearity of duplication genes were analyzed in *Gossypium* species including *G. hirsutum* and *G. arboreum*, *G. hirsutum*, and *G. raimondii*; *G. barbadense* and *G. arboreum*; *G. barbadense* and *G. raimondii*; and *G. hirsutum* and *G. barbadense* using the modified MCScan algorithm of the MCScanX package (default parameters) (<https://chibba.pgml.uga.edu/mcscan2/>) [61]. All results were drawn using Circos (<https://circos.ca/>) [62].

Analysis of chromosome location and gene regulatory elements

The chromosomal positions of all VQ members were determined using the gene transfer (gtf) format files of the reference genomes. The exon/intron structure of VQs were also extracted from the gtf files and displayed by the GSDS platform (<http://gsds.cbi.pku.edu.cn/>) [63]. Then, the MEME tool (<http://meme-suite.org/>) (the motif with 10 amino acids in length and E-value less than $1e^{-40}$) [64] was used to detect the additional motifs of the proteins. With the combination of TBtools software (v 0.67361) [65], all domain motifs were compared among VQ genes to identify the group-specific signatures.

Approximately 2000 bp genomic sequences locating upstream (from -2000 to -1) of the VQs from start codons were extracted from the cotton genome, which was

subsequently submitted to the online PlantCARE (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) [66] to determine the distribution of plant cis-acting regulatory elements. Moreover, the miRNA targets of the VQ members were predicted using the coding sequences (CDS) regions for complementary sequences by the psRNATarget server (<http://plantgrn.noble.org/psRNATarget/analysis?function=2>) [67] with default parameters, except maximum expectation (E) = 3.5. A total of 80 published miRNAs of *G. hirsutum* were selected.

Expression profiles and co-expression networks

To determine the expression patterns of VQs in the two allotetraploid kinds of *Gossypium*, RNA-Seq data were obtained from the SRA database (PRJNA490626) [33], including those in 10 tissues (the petal, pistil, root, sepal, stem, torus, filament, leaf, anther, and bract), during ovule development (-3, -1, 0, 1, 3, 5, 10, 20, and 25 DPA ovule) and fiber development (10, 20, and 25 DPA), and under four different abiotic stresses (NaCl, PEG, cold, and hot), which were previously generated and analyzed by Hu et al. [33]. Raw RNA-seq reads were filtered using the SRAToolkit (v 2.9.2) / fastq-dump [68] and trimmed by Trimmomatic (v 0.3.9) [69] to generate clean reads. Then, the filtered clean RNA-seq reads were mapped to their respective reference genomes using the HISAT2 (v 2.1.0) [70], and the Sequence Alignment/Map format (sam) data was converted to Binary Alignment/Map (bam) data using the samtools (v 1.9) [71]. The transcript abundances of annotated genes were quantified using the stringtie (v 2.0) [72]. The gene expression data were obtained from the big databases by R base, and expression levels expressed as log₂ (FPKM+1) and visualized using the pheatmap (v 1.0.12) [73]. The WGCNA (v 1.69) [74] and the STRING software (Searched Tool for the Retrieval of Interacting Genes/Proteins, <http://string-db.org/>) [75] were used for the analysis the co-expression networks, and the obtained results were displayed using the Cytoscape software (v 3.7.2) (<http://www.cytoscape.org/>) [76].

Plant cultivation, RNA isolation, and RT-PCR analysis

The germinated TM-1 cotton seeds were grown in plastic pots filled with the mixture of soil vermiculite, and the artificial growth conditions were set at 28/22 °C, with a photoperiod of 16 h light/8 h darkness. Plants were separately subjected to 400 mM PEG or 400 mM NaCl. Three biological replicates were sampled at 0, 1, 3, 6, 12, 24, 36, or 72 h. All the samples were collected and frozen in liquid nitrogen, which was stored at -80 °C until total RNA extraction.

Total RNAs of the above samples were isolated using the RNA prep Pure Plant Kit (Polysaccharides & Polyphenolics-rich, DP441) (TIANGEN, Beijing, China).

The concentrations and integrities of the extracted RNA samples were measured and verified using a NanoDrop machine and 1% agarose gel electrophoresis, and the RNA samples were reversed transcribed into complementary DNA (cDNA) using the Mir-X™ MIRNA First-Strand Synthesis Kit (TaKaRa, Dalian, China). The qRT-PCR was performed using the Roche LightCycler 480 System (Roche, Germany). The qRT-PCR primers for the *GhVQs* and actin gene were listed in Supplemental Table S8. The reaction was set up in a total volume of 20 μ L: 2 μ L (200 ng) of cDNA, 0.4 μ L of forwarding primer (10.0 μ M), 0.4 μ L of reverse primer (10.0 μ M), 10 μ L 2 \times TransStart Top/Tip Green qPCR SuperMix, and 7.2 μ L of nuclease-free water. The reaction procedure was completed under the following program: 94 °C for 30 s; 45 cycles of 94 °C for 5 s, 60 °C for 15 s, 72 °C for 10 s; and 4 °C to finish. The results were calculated using the $2^{-\Delta\Delta C_t}$ relative quantitative method.

Supplementary Information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12864-020-07171-z>.

Additional file 1: Table S1. Summaries of VQ genes members in *Gossypium*.

Additional file 2: Figure S1. The physical and chemical properties of VQs in *Gossypium spp.* with violin illustration. **a, b, c, d, e, f** are the lengths of the transcripts, GC contents, exon numbers, protein lengths, PI values and MW values, respectively.

Additional file 3: Table S2. The number of VQ genes in the 15 plant genomes.

Additional file 4: Figure S2. The phylogenetic tree of VQs in *Gossypium spp.* compared with *A. thaliana*. **a** The phylogenetic tree of VQs in *G. hirsutum* and *A. thaliana*. **b** The phylogenetic tree of VQs in *G. barbadense* and *A. thaliana*. **c** The phylogenetic tree of VQs in *G. raimondii* and *A. thaliana*. **d** The phylogenetic tree of VQs in *G. arboretum* and *A. thaliana*.

Additional file 5: Figure S3. The conserved motifs of VQs.

Additional file 6: Figure S4. The paralogs of VQs between *G. hirsutum* and another two diploid *Gossypium* species. The lines regarding orthologous gene pairs are colored by green and blue. The green lines are the pairs between *G. hirsutum* D-subgenome and *G. raimondii* genome, and the blue lines are the pairs between *G. hirsutum* A-subgenome and *G. arboretum* genome.

Additional file 7: Table S3. The list of paralogous VQ gene pairs in each of other examined species.

Additional file 8: Figure S5. The paralogs of VQs between *G. barbadense* and another two diploid *Gossypium* species. The lines regarding orthologous gene pairs are colored by green and blue. The green lines are the pairs between *G. barbadense* D-subgenome and *G. raimondii* genome; and the blue lines are the pairs between *G. barbadense* A-subgenome and *G. arboretum* genome.

Additional file 9: Table S4. The duplicated type of VQs in *Gossypium spp.* The 0 to 4 indicate the singleton, dispersed, proximal, tandem, WGD duplication types, respectively.

Additional file 10: Table S5. List of predicted known microRNA target sites of VQ transcripts in *Gossypium* species.

Additional file 11: Figure S6. Expression patterns of *GhVQs* in different tissues and under stresses in *G. hirsutum*.

Additional file 12: Figure S7. Expression patterns of *GbVQs* in different tissues and under stresses in *G. barbadense*.

Additional file 13: Table S6. Expression levels of *GhVQs* and *GbVQs*. The gene expression was determined by RNA-Seq data (FPKM).

Additional file 14: Table S7. Information regarding the Co-expression and STRING search results of the interaction networks of the *GhVQs*.

Additional file 15: Table S8. Primers used for qPCR analyses of *GhVQs*.

Acknowledgements

We are grateful to all the colleagues in our laboratory.

Authors' contributions

S.X.Y. and H.L.W. designed the research program. F.W., S.S.C., L.M. and H.T.W. analyzed and transcriptomic data. M.Z., G.Z.M., J.H.L., P.B.H., and A.M.W. participated in cotton culture processing and RNA extraction, A.A., L.J.G., and Q.M. revised the language. P.Y.C. performed the experiment and wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This project is based on research that was supported by Central Public-interest Scientific Institution Basal Research Fund (No. Y2018YJ05).

Availability of data and materials

All data supporting the conclusions of this article are included in the article and its additional files.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹State Key Laboratory of Cotton Biology, Institute of Cotton Research of Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, China.

²School of Life Science, Zhengzhou University, Zhengzhou 450000, China.

³College of Agronomy, Northwest A&F University, Yangling 712100, China.

Received: 23 June 2020 Accepted: 21 October 2020

Published online: 16 November 2020

References

- Wang A, Garcia D, Zhang H, Feng K, Chaudhury A, Berger F, Peacock WJ, Dennis ES, Luo M. The VQ motif protein IKU1 regulates endosperm growth and seed size in Arabidopsis. *Plant J.* 2010;63(4):670–9.
- Jiang SY, Sevugan M, Ramachandran S. Valine-glutamine (VQ) motif coding genes are ancient and non-plant-specific with comprehensive expression regulation by various biotic and abiotic stresses. *BMC Genomics.* 2018;19(1):342.
- Cai HY, Zhang M, Liu YH, He Q, Chai MN, Liu LP, Chen FQ, Huang YM, Yan MK, Zhao HM, et al. Genome-wide classification and evolutionary and functional analyses of the VQ family. *Trop Plant Biol.* 2019;12(2):117–31.
- Xie YD, Li W, Guo D, Dong J, Zhang Q, Fu Y, Ren D, Peng M, Xia Y. The *Arabidopsis* gene sigma factor-binding protein 1 plays a role in the salicylate- and jasmonate-mediated defence responses. *Plant Cell Environ.* 2010;33(5):828–39.
- Jing Y, Lin R. The VQ motif-containing protein family of plant-specific transcriptional regulators. *Plant Physiol.* 2015;169(1):371–8.
- Lai Z, Li Y, Wang F, Cheng Y, Fan B, Yu JQ, Chen Z. Arabidopsis sigma factor binding proteins are activators of the WRKY33 transcription factor in plant defense. *Plant Cell.* 2011;23(10):3824–41.
- Jiang Y, Yu D. The WRKY57 transcription factor affects the expression of Jasmonate ZIM-domain genes transcriptionally to compromise Botrytis cinerea resistance. *Plant Physiol.* 2016;171(4):2771–82.
- Hu P, Zhou W, Cheng Z, Fan M, Wang L, Xie D. JAV1 controls jasmonate-regulated plant defense. *Mol Cell.* 2013;50(4):504–15.

9. Hu Y, Chen L, Wang H, Zhang L, Wang F, Yu D. Arabidopsis transcription factor WRKY8 functions antagonistically with its interacting partner VQ9 to modulate salinity stress tolerance. *Plant J.* 2013;74(5):730–45.
10. Dong Q, Zhao S, Duan D, Tian Y, Wang Y, Mao K, Zhou Z, Ma F. Structural and functional analyses of genes encoding VQ proteins in apple. *Plant Sci.* 2018;272:208–19.
11. Xixu P, Ting X, Jiao M, Zong T, Ding-gang Z, Xin-ke T, Hai-hua W. Differential expression of Rice Valine- α -Glutamine gene family in response to nitric oxide and regulatory circuit of OsVQ7 and OsWRKY24. *Rice Sci.* 2019;27:10–20.
12. Lei R, Li X, Ma Z, Lv Y, Hu Y, Yu D. Arabidopsis WRKY 2 and WRKY 34 transcription factors interact with VQ 20 protein to modulate pollen development and function. *Plant J.* 2017;91(6):962–76.
13. Pecher P, Eschen-Lippold L, Herklotz S, Kuhle K, Naumann K, Bethke G, Uhrig J, Weyhe M, Scheel D, Lee J. The Arabidopsis thaliana mitogen-activated protein kinases MPK3 and MPK6 target a subclass of 'VQ-motif'-containing proteins to regulate immune responses. *New Phytol.* 2014;203(2):592–606.
14. Xie F, Wang Q, Sun R, Zhang B. Deep sequencing reveals important roles of microRNAs in response to drought and salinity stress in cotton. *J Exp Bot.* 2015;66(3):789–804.
15. Eulgem T, Rushton PJ, Robatzek S, Somssich IE. The WRKY superfamily of plant transcription factors. *Trends Plant Sci.* 2000;5(5):199–206.
16. Dou L, Zhang X, Pang C, Song M, Wei H, Fan S, Yu S. Genome-wide analysis of the WRKY gene family in cotton. *Mol Gen Genomics.* 2014;289(6):1103–21.
17. Grandbastien MA. LTR retrotransposons, handy hitchhikers of plant regulation and stress response. *Biochim Biophys Acta.* 2015;1849(4):403–16.
18. Zhu JK. Plant salt tolerance. *Trends Plant Sci.* 2001;6(2):66–71.
19. Liu C, Liu H, Zhou C, Timko MP. Genome-Wide Identification of the VQ Protein Gene Family of Tobacco (*Nicotiana tabacum* L.) and Analysis of Its Expression in Response to Phytohormones and Abiotic and Biotic Stresses. *Genes (Basel).* 2020;11(3):284.
20. Roberts EM, Rao NR, Huang JY, Trolinder NL, Haigler CH. Effects of cycling temperatures on fiber metabolism in cultured cotton ovules. *Plant Physiol.* 1992;100(2):979–86.
21. Khan A, Pan X, Najeeb U, Tan DKY, Fahad S, Zahoor R, Luo H. Coping with drought: stress and adaptive mechanisms, and management through cultural and molecular alternatives in cotton as vital constituents for plant stress resilience and fitness. *Biol Res.* 2018;51(1):47.
22. Wang W, Vinocur B, Altman A. Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta.* 2003;218(1):1–14.
23. Cramer GR, Urano K, Delrot S, Pezzotti M, Shinozaki K. Effects of abiotic stress on plants: a systems biology perspective. *BMC Plant Biol.* 2011; 11(1):163.
24. Li Y, Jing Y, Li J, Xu G, Lin R. Arabidopsis VQ motif-containing protein29 represses seedling deetiolation by interacting with phytochrome-interacting factor1. *Plant Physiol.* 2014;164(4):2068–80.
25. Wang Y, Jiang Z, Li Z, Zhao Y, Tan W, Liu Z, Cui S, Yu X, Ma J, Wang G, et al. Genome-wide identification and expression analysis of the VQ gene family in soybean (*Glycine max*). *PeerJ.* 2019;7:e7509.
26. Kim DY, Kwon SI, Choi C, Lee H, Ahn I, Park SR, Bae SC, Lee SC, Hwang DJ. Expression analysis of rice VQ genes in response to biotic and abiotic stresses. *Gene.* 2013;529(2):208–14.
27. Song W, Zhao H, Zhang X, Lei L, Lai J. Genome-wide identification of VQ motif-containing proteins and their expression profiles under abiotic stresses in maize. *Front Plant Sci.* 2016;6:1177.
28. Zhang G, Wang F, Li J, Ding Q, Zhang Y, Li H, Zhang J, Gao J. Genome-wide identification and analysis of the VQ motif-containing protein family in Chinese cabbage (*Brassica rapa* L. ssp. *pekinensis*). *Int J Mol Sci.* 2015;16(12): 28683–704.
29. Guo J, Chen J, Yang J, Yu Y, Yang Y, Wang W. Identification, characterization and expression analysis of the VQ motif-containing gene family in tea plant (*Camellia sinensis*). *BMC Genomics.* 2018;19(1):710.
30. Ding H, Yuan G, Mo S, Qian Y, Wu Y, Chen Q, Xu X, Wu X, Ge C. Genome-wide analysis of the plant-specific VQ motif-containing proteins in tomato (*Solanum lycopersicum*) and characterization of SIVQ6 in thermotolerance. *Plant Physiol Biochem.* 2019;143:29–39.
31. Ma J, Ling L, Huang X, Wang W, Yb W, Zhang M, Sy C, Zhou F, Yk Q, Cb L, et al. Genome-wide identification and expression analysis of the VQ gene family in sunflower (*Helianthus annuus* L.). *J Plant Biochem Biotechnol.* 2020:1–11. <https://doi.org/10.1007/s13562-020-00568-7>.
32. Ling L, Qu Y, Zhu J, Wang D, Guo C. Genome-wide identification and expression analysis of the VQ gene family in *Cicer arietinum* and *Medicago truncatula*. *PeerJ.* 2020;8:e8471.
33. Hu Y, Chen J, Fang L, Zhang Z, Ma W, Niu Y, Ju L, Deng J, Zhao T, Lian J, 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.
34. Du X, Huang G, He S, Yang Z, Sun G, Ma X, Li N, Zhang X, Sun J, Liu M, 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.
35. Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, et al. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature.* 2012;492(7429):423–7.
36. Lee TH, Tang H, Wang X, Paterson AH. PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res.* 2012;41(D1):D1152–8.
37. Wang H, Hu Y, Pan J, Yu D. Arabidopsis VQ motif-containing proteins VQ12 and VQ29 negatively modulate basal defense against *Botrytis cinerea*. *Sci Rep.* 2015;5(1):14185.
38. Pan JJ, Wang HP, Hu Y, Yu D. Arabidopsis VQ18 and VQ26 proteins interact with AB15 transcription factor to negatively modulate ABA response during seed germination. *Plant J.* 2018;95(3):529–44.
39. Zhang Y, Zhu X, Chen X, Song C, Zou Z, Wang Y, Wang M, Fang W, Li X. Identification and characterization of cold-responsive microRNAs in tea plant (*Camellia sinensis*) and their targets using high-throughput sequencing and degradome analysis. *BMC Plant Biol.* 2014;14(1):271.
40. Wu G, Park MY, Conway SR, Wang JW, Weigel D, Poethig RS. The sequential action of miR156 and miR172 regulates developmental timing in *Arabidopsis*. *Cell.* 2009;138(4):750–9.
41. Naqvi AR, Haq QM, Mukherjee SK. MicroRNA profiling of tomato leaf curl New Delhi virus (toLCNDV) infected tomato leaves indicates that deregulation of miR159/319 and miR172 might be linked with leaf curl disease. *Virology.* 2010;7(1):281.
42. Stief A, Altmann S, Hoffmann K, Pant BD, Scheible WR, Baurle I. Arabidopsis miR156 regulates tolerance to recurring environmental stress through SPL transcription factors. *Plant Cell.* 2014;26(4):1792–807.
43. Ye YJ, Xiao YY, Han YC, Shan W, Fan ZQ, Xu QG, Kuang JF, Lu WJ, Lakshmanan P, Chen JY. Banana fruit VQ motif-containing protein5 represses cold-responsive transcription factor MaWRKY26 involved in the regulation of JA biosynthetic genes. *Sci Rep.* 2016;6(1):23632.
44. Wang Y, Liu H, Zhu D, Gao Y, Yan H, Xiang Y. Genome-wide analysis of VQ motif-containing proteins in Moso bamboo (*Phyllostachys edulis*). *Planta.* 2017;246(1):165–81.
45. Adams KL, Wendel JF. Polyploidy and genome evolution in plants. *Curr Opin Plant Biol.* 2005;8(2):135–41.
46. Fawcett JA, Maere S, Van de Peer Y. Plants with double genomes might have had a better chance to survive the cretaceous-tertiary extinction event. *Proc Natl Acad Sci U S A.* 2009;106(14):5737–42.
47. Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. Genome size diversity and its impact on the evolution of land plants. *Genes (Basel).* 2018;9(2):88.
48. Wang M, Vannozzi A, Wang G, Zhong Y, Corso M, Cavallini E, Cheng ZM. A comprehensive survey of the grapevine VQ gene family and its transcriptional correlation with WRKY proteins. *Front Plant Sci.* 2015;6:417.
49. Cheng Y, Zhou Y, Yang Y, Chi YJ, Zhou J, Chen JY, Wang F, Fan B, Shi K, Zhou YH, et al. Structural and functional analysis of VQ motif-containing proteins in *Arabidopsis* as interacting proteins of WRKY transcription factors. *Plant Physiol.* 2012;159(2):810–25.
50. Perruc E, Charpentreau M, Ramirez BC, Jauneau A, Galaud JP, Ranjeva R, Ranty B. A novel calmodulin-binding protein functions as a negative regulator of osmotic stress tolerance in *Arabidopsis thaliana* seedlings. *Plant J.* 2004;38(3):410–20.
51. Kumar S, Stecher G, Suleski M, Hedges SB. TimeTree: a resource for timelines, Timetrees, and divergence times. *Mol Biol Evol.* 2017;34(7):1812–9.
52. Altschul SF. Basic local alignment search tool (BLAST). *J Mol Biol.* 1990; 215(3):403–10.
53. Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, Carol D-S, Fong JH, Geer LY, Geer RC, Gonzales NR. CDD: a conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* 2010;39:225–9.
54. Ivica L, Tobias D, Peer B. SMART 6: recent updates and new developments. *Nucleic Acids Res.* 2009;37:229–32.

55. Panu A, Manohar J, Konstantin A, Delphine B, Gabor C, Edouard DC, Séverine D, Volker F, Arnaud F, Elisabeth G. ExpASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* 2012;40:597–603.
56. Wickham H. ggplot2: elegant graphics for data analysis. New York: Springer; 2009. <http://had.co.nz/ggplot2/book>.
57. Katoh K, Standley DM. MAFFT: iterative refinement and additional methods. *Methods Mol Biol.* 2014;1079(1079):131–46.
58. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56(4):564–77.
59. Nguyen L, Schmidt H, Haeseler AV, Minh B. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;1:268–74.
60. Yu GC, Smith DK, Hc Z, Guan Y, Lam TT, McInerney G. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 2016;8(1):28–36.
61. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, et al. MCS-X: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 2012;40(7):e49.
62. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–45.
63. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics.* 2015;31(8):1296–7.
64. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009;37:W202–8.
65. Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R. TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Mol Plant.* 2020;13:1194–202.
66. Lescot M, Dehais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouze P, Rombauts S. 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.
67. Dai X, Zhao PX. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Res.* 2011;39:W155–9.
68. Leinonen R, Sugawara H, Shumway M. International nucleotide sequence database C: the sequence read archive. *Nucleic Acids Res.* 2011;39(Database issue):D19–21.
69. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30(15):2114–20.
70. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* 2015;12(4):357–60.
71. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing S: the sequence alignment/map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–9.
72. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol.* 2015;33(3):290–5.
73. Kolde R, Kolde MR: Package ‘pheatmap’. Bioconductor 1–6. Available at: <https://CRAN.R-project.org/package=pheatmap>. 2012.
74. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 2008;9(1):559.
75. Wagner R, Fischer M. The string-to-string correction problem. *J ACM (JACM).* 1974;21(1):168–73.
76. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–504.

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

