BMC Genomics

Genome-wide characterization of the GRF family and their roles in response to salt stress in *Gossypium*

Jun-Feng Cao[1,2,3], Jin-Quan Huang[1], Xia Liu[4], Chao-Chen Huang[1,5], Zi-Shou Zheng[1,3], Xiu-Fang Zhang[1], Xiao-Xia Shangguan[1], Ling-Jian Wang[1], Yu-Gao Zhang[4], Jonathan F. Wendel[6], Corrinne E. Grover[6*] and Zhi-Wen Chen[1,7*]

Abstract

Background: Cotton (*Gossypium* spp.) is the most important world-wide fiber crop but salt stress limits cotton production in coastal and other areas. Growth regulation factors (GRFs) play regulatory roles in response to salt stress, but their roles have not been studied in cotton under salt stress.

Results: We identified 19 GRF genes in *G. raimondii*, 18 in *G. arboreum*, 34 in *G. hirsutum* and 45 in *G. barbadense*, respectively. These GRF genes were phylogenetically analyzed leading to the recognition of seven GRF clades. GRF genes from diploid cottons (*G. raimondii* and *G. arboreum*) were largely retained in allopolyploid cotton, with subsequent gene expansion in *G. barbadense* relative to *G. hirsutum*. Most *G. hirsutum* GRF (*GhGRF*) genes are preferentially expressed in young and growing tissues. To explore their possible role in salt stress, we used qRT-PCR to study expression responses to NaCl treatment, showing that five *GhGRF* genes were down-regulated in leaves. RNA-seq experiments showed that seven *GhGRF* genes exhibited decreased expression in leaves under NaCl treatment, three of which (*GhGRF3*, *GhGRF4*, and *GhGRF16*) were identified by both RNA-seq and qRT-PCR. We also identified six and three *GRF* genes that exhibit decreased expression under salt stress in *G. arboreum* and *G. barbadense*, respectively. Consistent with its lack of leaf withering or yellowing under the salt treatment conditions, *G. arboreum* had better salt tolerance than *G. hirsutum* and *G. barbadense*. Our results suggest that GRF genes are involved in salt stress responses in *Gossypium*.

Conclusion: In summary, we identified candidate GRF genes that were involved in salt stress responses in cotton.

Keywords: Growth regulating factor, Cotton, Phylogenetics, Organ-specific expression, Salt stress

* Correspondence: corrinne@iastate.edu; b1301031@cau.edu.cn
[6]Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA
[1]National Key Laboratory of Plant Molecular Genetics and National Center for Plant Gene Research, Institute of Plant Physiology and Ecology/CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China
Full list of author information is available at the end of the article

Cao *et al. BMC Genomics*   (2020) 21:575

Page 2 of 16

Background

Cotton (from the genus *Gossypium*) is the most important fiber crop in the world. While the genus itself contains more than 50 species, cultivated cottons are composed of four independently domesticated species, among which the Upland cotton (*G. hirsutum* L.) is the most prominent cultivar, accounting for 95% of global cotton fiber output [1, 2]. Due to the agronomic importance of cotton, it is worth studying its defense and stress responses. Cotton plants produce gossypol and related sesquiterpene aldehydes that function as defensive compounds against pests and pathogens [3–7]. With respect to abiotic stresses, cotton has a moderate tolerance to salinity and consequently is cultivated in some saline-alkali soils; however, salt stress is still a limiting factor that affects cotton production [8].

Identification and characterization of salt response genes is important in dissecting the molecular mechanisms of plant adaptation to salt stresses and, ultimately, in engineering salt-tolerant crops. Consequently, research into the genetic responses to salt stress has yielded much information (see recent reviews, such as [9–12]), including the role of growth regulation on salt tolerance. Growth-regulating factors (GRFs) are plant-specific transcription factors, which form a relatively small family and function in plant development and stress responses [13]. The first GRF was identified from rice, which uncovered a regulatory role for GRFs during leaf and stem development [14]. Subsequent research showed that these transcription factors are also involved in other aspects of plant growth and adaptation, including root development [15], flowering [16, 17], leaf size and longevity [18], and response to abiotic stresses [19–22]. Recent research on abiotic stress responses have reported a role for GRF repression in abiotic stress tolerance. For example, the activation of the *Arabidopsis* stress-responsive gene AtDREB2A, whose expression increases plant tolerance to osmotic stress [23, 24], requires repression of at least one GRF. Under non-stress conditions, AtGRF7 binds to the *DREB2A* promoter to suppress expression; however, abiotic stress leads to suppression of AtGRF7 expression and consequently the activation of osmotic stress-responsive genes [20].

Recent advances in cotton genomics have produced the resources necessary to characterize the GRF gene family in *Gossypium.* Multiple high-quality genome sequences are available for several species, including two diploid species, i.e., *G. raimondii* (D_5) [25, 26] and *G. arboreum* (A_2; cultivated) [27, 28], and two cultivated allotetraploids, namely *G. hirsutum* (AD_1) [29–31] and *G. barbadense* (AD_2) [31–34]. These resources provide the foundation for identifying the suite of GRF genes in *Gossypium* and their potential roles as salt stress-related genes. Here, we report our genome-wide analysis of the GRF transcription factor genes in four important cotton species, reporting a modestly sized family (18 and 19 members in diploid species, 32 and 45 in polyploids) that are largely conserved during evolution and polyploidization. We further explore the role of GRF genes in response to salt stress. We identified three *G. hirsutum* GRF genes, six *G. arboreum* GRF genes, and three *G. barbadense* GRF genes that exhibit decreased expression under salt stress, respectively. Compared to the tetraploid species, diploid cotton *G. arboreum* was more salt tolerant. These candidate GRF genes may be useful in future molecular breeding of salt-tolerant cotton species.
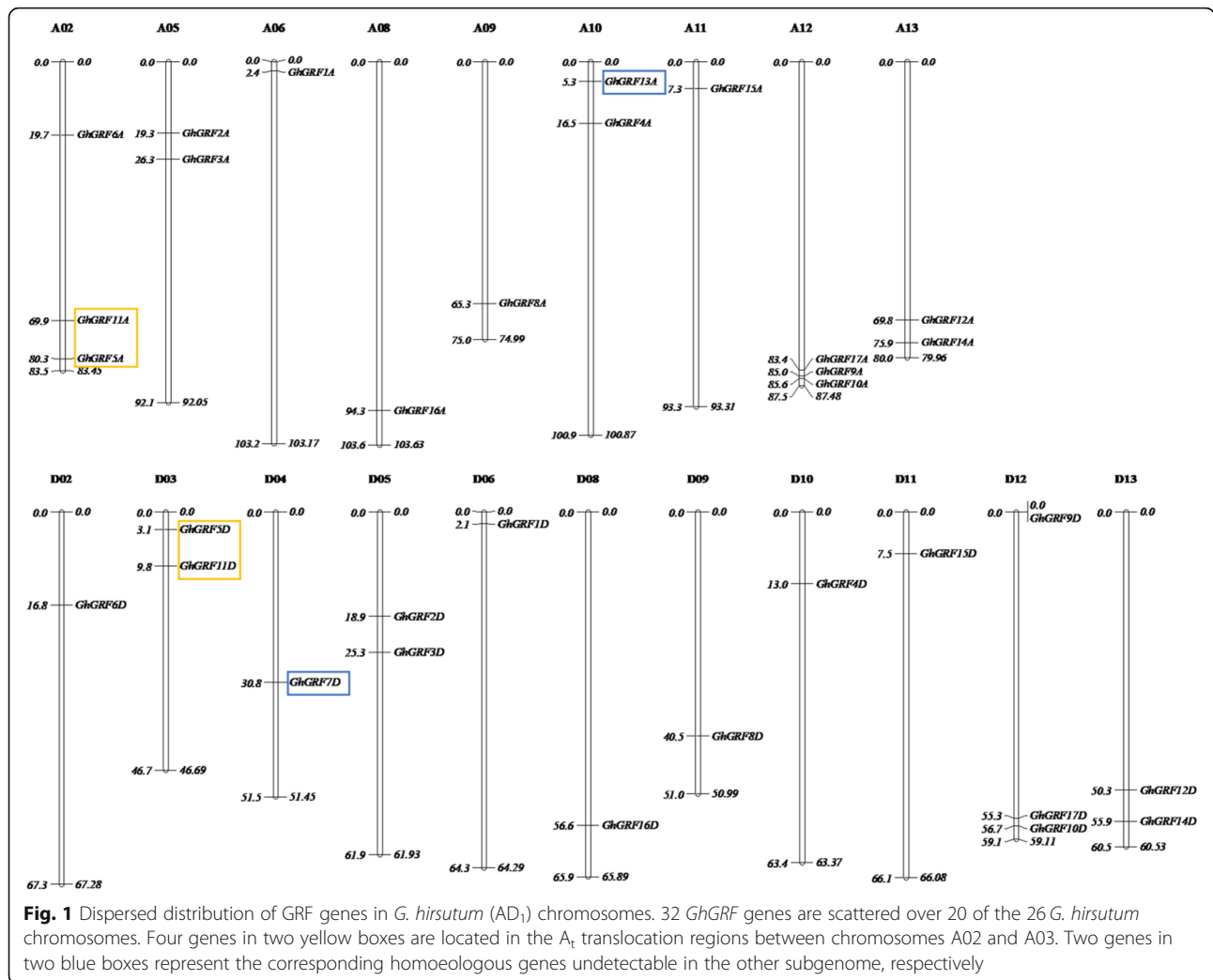
Results

Genome-wide identification and sequence analysis of genes encoding putative GRFs in *G. hirsutum*

GRF proteins are involved in plant responses to abiotic stresses [19, 22], including salt stress [20]. These proteins are defined by the presence of a QLQ domain for protein-protein interactions, a WRC domain responsible for DNA binding and a putative nuclear localization signal [13, 14, 35]. Hmmersearch against the *G. hirsutum* genome database with these two conserved domains (PF08879 for WRC domain and PF08880 for QLQ domain) identified 34 putative GRF-encoding genes, here designated *GhGRF1A–GhGRF17A* (A homoeologs) and *GhGRF1D–GhGRF17D* (D homoeologs), where the A and D suffix in the gene name indicates the genome of origin. These 34 *GhGRF* genes are dispersed over 20 of the 26 *G. hirsutum* chromosomes, with most, but not all, homoeologs conserved (Fig. 1). Syntenic conservation of two homoeologous pairs, i.e., *GhGRF5A/D* and *GhGRF11A/D*, was disrupted by a known large chromosomal translocation [29, 30, 32]. This translocation resulted in these conserved homoeologs being located on the non-homoeologous chromosomes A02 and D03 (Fig. 1). In addition, a single homoeolog each was recovered for both *GhGRF7D* and *GhGRF13A* (Fig. 1) by conserved domain search; however, BLAST recovered syntenic copies of GRF-like genes for both genes that could represent the missing homoeologs, i.e., *GhGRF7A* and *GhGRF13D*. Both syntenic GRF-like homoeologs lack the canonical conserved GRF protein domains, possibly indicating a loss of, or change in, function; therefore, these homoeologs were excluded from further analyses.

Structural organization of *GhGRF* genes

More than threefold variation in length was detected in the predicted coding sequences (CDS) for the recovered *GhGRF*s, from 546 bp for *GhGRF14A/D* to 1833 bp for *GhGRF2A/D*, (Table 1), which translates to proteins ranging from 181 amino acids (aa) (20.73 kDa) to 610 aa (65.58 kDa). Predicted isoelectric points (pI) for members of this family also vary widely, from 5.89 to 9.59,

Cao et al. BMC Genomics (2020) 21:575

Page 3 of 16

Fig. 1 Dispersed distribution of GRF genes in *G. hirsutum* (AD$_1$) chromosomes. 32 *GhGRF* genes are scattered over 20 of the 26 *G. hirsutum* chromosomes. Four genes in two yellow boxes are located in the A$_t$ translocation regions between chromosomes A02 and A03. Two genes in two blue boxes represent the corresponding homoeologous genes undetectable in the other subgenome, respectively

possibly due to the composition of the C-terminal region (Additional file 3: Figure S1). All of the putative GhGRF proteins have both QLQ and WRC domains [14] in the N-terminal region (Table 1, Fig. 2, Additional file 3: Figure S1). GhGRF15A and GhGRF15D also contain a second WRC domain downstream of the first one (Additional file 3: Fig. S1). A zinc finger motif (CCCH) was also observed within the WRC domain in all putative GhGRF proteins (Fig. 2).

While all putative *GhGRF* genes contain predicted introns (Fig. 3), they also exhibit considerable variation, in both length and number. In general, homoeologous *GRF* genes show highly similar intron patterns (see exceptions below); however, intron structure among homoeologous pairs can exhibit variation in intron number (1 to 4) and length. Three of the homoeologous gene pairs did exhibit divergence in structure between homoeologs, namely *GhGRF9A* vs *GhGRF9D*, *GhGRF16A* vs *GhGRF16D* and *GhGRF17A* vs *GhGRF17D*. In two of the three cases (i.e., *GhGRF9*, *GhGRF16*), the acquisition of a splice site in the

A-homoeolog led to an additional, large intron. For *GhGRF17A/D*, however, the D-homoeolog exhibits loss of splicing for the first intron, resulting in read-through transcription. Characterization of parental (in the diploids) gene structure for these three homoeologs suggests that this structural variation represents inherited, parental divergence. Phylogenetic analysis of the GRF gene family combined with intron/exon structure characterization naturally divides the family into six groups, here designated A-F, containing 3 to 12 genes (Fig. 3).

Phylogenetic analysis of GRF proteins in *Gossypium*

The general conservation of GRF genes between the two homoeologous genomes of *G. hirsutum* suggests minimal loss and/or gain since the divergence of the progenitor diploid genomes. We specifically assessed this using the protein sequences of 114 cotton GRFs (*G. hirsutum*, 32; *G. barbadense*, 45; *G. raimondii*, 19; and *G. arboreum*, 18; see methods) with 9 *Arabidopsis thaliana* GRFs for phylogenetic analysis (Fig. 4). The six

Table 1 Sequence characteristics of *GhGRF* (*Gossypium hirsutum* growth-regulating factor) genes and corresponding proteins

Gene Name	Locus Name	Chr	Genomics position	CDS	No. of Introns	Size (aa)	QLQ Domain	WRC Domain	MW	pI
GhGRF1A	Gh_A06G0219	A06	2,427,674–2,430,873	1782	3	593	161–197	232–274	64.42	8.83
GhGRF1D	Gh_D06G0212	D06	2,081,549–2,084,687	1731	3	576	162–198	233–275	62.37	6.48
GhGRF2A	Gh_A05G1848	A05	19,322,416–19,325,837	1833	3	610	162–198	231–273	65.38	8.36
GhGRF2D	Gh_D05G2044	D05	18,868,989–18,872,409	1833	3	610	160–196	231–273	65.58	8.06
GhGRF3A	Gh_A05G2252	A05	26,265,366–26,267,943	1815	3	604	143–179	215–257	65.27	6.58
GhGRF3D	Gh_D05G2512	D05	25,255,757–25,258,332	1815	3	604	143–179	215–257	65.41	6.48
GhGRF4A	Gh_A10G0804	A10	16,451,418–16,453,809	1428	3	475	139–175	210–252	52.25	6.63
GhGRF4D	Gh_D10G0959	D10	12,983,506–12,985,893	1428	3	475	139–175	210–252	52.03	6.34
GhGRF5A	Gh_A02G1438	A02	80,257,374–80,259,722	1569	3	522	132–168	203–245	57.25	7.71
GhGRF5D	Gh_D03G0282	D03	3,085,927–3,089,051	1686	3	561	140–176	211–253	61.24	8.14
GhGRF6A	Gh_A02G0827	A02	19,707,965–19,710,598	1695	3	564	144–180	216–258	61.30	7.26
GhGRF6D	Gh_D02G0876	D02	16810,024–16,812,651	1710	3	569	144–180	216–258	61.70	6.97
GhGRF7D	Gh_D04G1000	D04	30,836,475–30,838,388	1092	3	363	73–109	147–189	39.96	9.13
GhGRF8A	Gh_A09G1285	A09	65,252,506–65,254,792	1218	3	405	83–119	149–191	44.21	8.25
GhGRF8D	Gh_D09G1332	D09	40,520,770–40,523,059	1209	3	402	81–117	147–189	43.77	8.54
GhGRF9A	Gh_A12G2234	A12	85,044,073–85,048,399	1491	4	496	101–137	158–200	53.51	6.96
GhGRF9D	Gh_D12G2699	D12	7626–9481	1491	3	496	102–138	159–201	53.58	6.94
GhGRF10A	Gh_A12G2177	A12	84,589,570–84,591,240	1401	3	466	152–188	210–252	50.86	5.89
GhGRF10D	Gh_D12G2356	D12	56,690,643–56,692,281	1377	3	458	144–180	202–244	50.05	5.95
GhGRF11A	Gh_A02G1205	A02	69,876,127–69,879,623	1008	2	335	30–66	93–135	36.71	7.74
GhGRF11D	Gh_D03G0527	D03	9,819,619–9,823,083	1026	2	341	30–66	93–135	37.40	7.81
GhGRF12A	Gh_A13G1365	A13	69,805,651–69,809,400	1005	2	334	30–66	93–135	36.42	8.30
GhGRF12D	Gh_D13G1673	D13	50,330,408–50,334,209	996	2	331	27–63	90–132	36.13	8.29
GhGRF13A	Gh_A10G0492	A10	5,324,275–5,325,500	987	2	328	8–44	75–117	36.51	9.02
GhGRF14A	Gh_A13G1692	A13	75,869,496–75,870,119	546	1	181	12–48	75–117	20.73	9.26
GhGRF14D	Gh_D13G2042	D13	55,933,517–55,934,140	546	1	181	12–48	75–117	20.99	9.48
GhGRF15A	Gh_A11G0749	A11	7,286,137–7,288,429	1338	3	445	70–104	141–183/337–378	48.82	9.59
GhGRF15D	Gh_D11G0870	D11	7,509,881–7,512,169	1338	3	445	70–104	141–183/337–378	48.63	9.58
GhGRF16A	Gh_A08G1584	A08	94,278,338–94,283,477	774	3	257	65–101	159–189	27.79	8.89
GhGRF16D	Gh_D08G1891	D08	56,643,392–56,644,288	744	2	247	65–101	137–179	26.71	8.89
GhGRF17A	Gh_A12G2042	A12	83,346,587–83,347,484	729	2	242	58–94	130–172	26.37	9.10
GhGRF17D	Gh_D12G2219	D12	55,327,073–55,327,959	804	1	267	83–119	155–197	29.24	9.14

Note: *bp* Base pair, *Chr* Chromosome, *aa* Amino acid, *MW* Molecular weight, *kDa* Kilodalton, *pI* Isoelectric point

Cao *et al. BMC Genomics*   (2020) 21:575

Page 4 of 16

Fig. 2 Sequence alignment of GhGRF (*Gossypium hirsutum* GRF) proteins and the QLQ and WRC domains are indicated upside. Identical amino acids are indicated by the color background
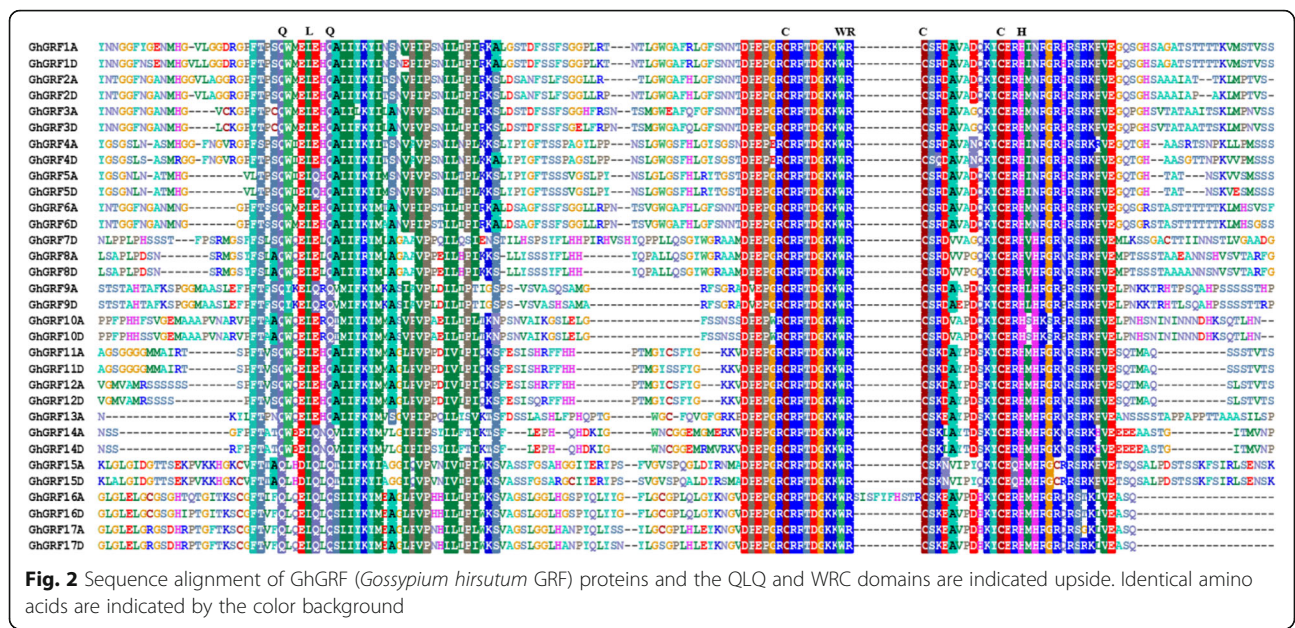
previously designated clades (A–F) were again recovered plus extra clade "G", with 1–2 *A. thaliana* genes associated with each clade except clades B and G, which are composed of *Gossypium* GRFs only.

Overall, the expected diploid-polyploid topology is reflected in the tree for each set of orthologous/ homoeologous genes, indicating general preservation during diploid divergence and through polyploid evolution. That is, the number of GRF genes in *G. hirsutum* was generally additive with respect to the model diploid progenitors (*G. raimondii* and *G. arboreum*), with each homoeolog (A_t or D_t) sister to their respective parental copies.
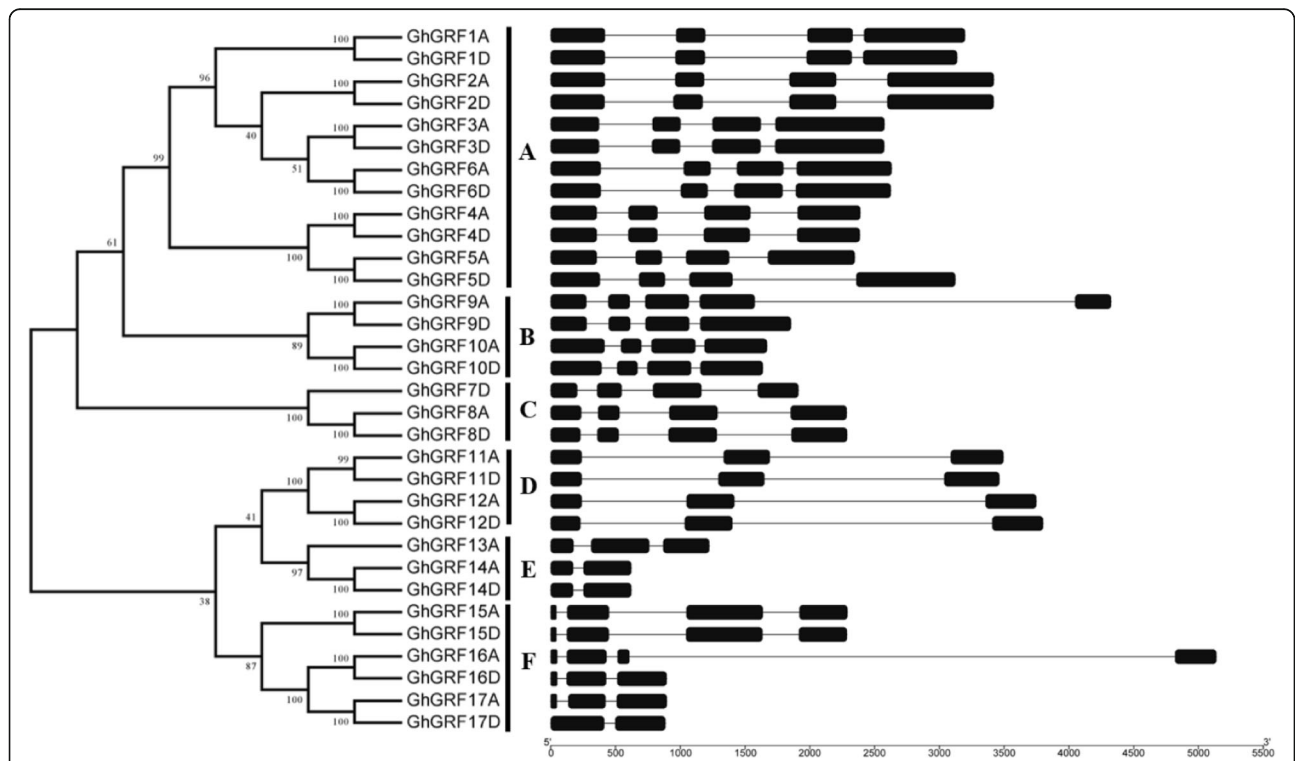

Fig. 3 Phylogenetic tree and gene structure of GRF protein genes in *G. hirsutum*. Six phylogenetic clades were clustered. Exons and introns are represented by black boxes and lines, respectively. **a-f** represent six phylogenetic clades with different intron/exon structure

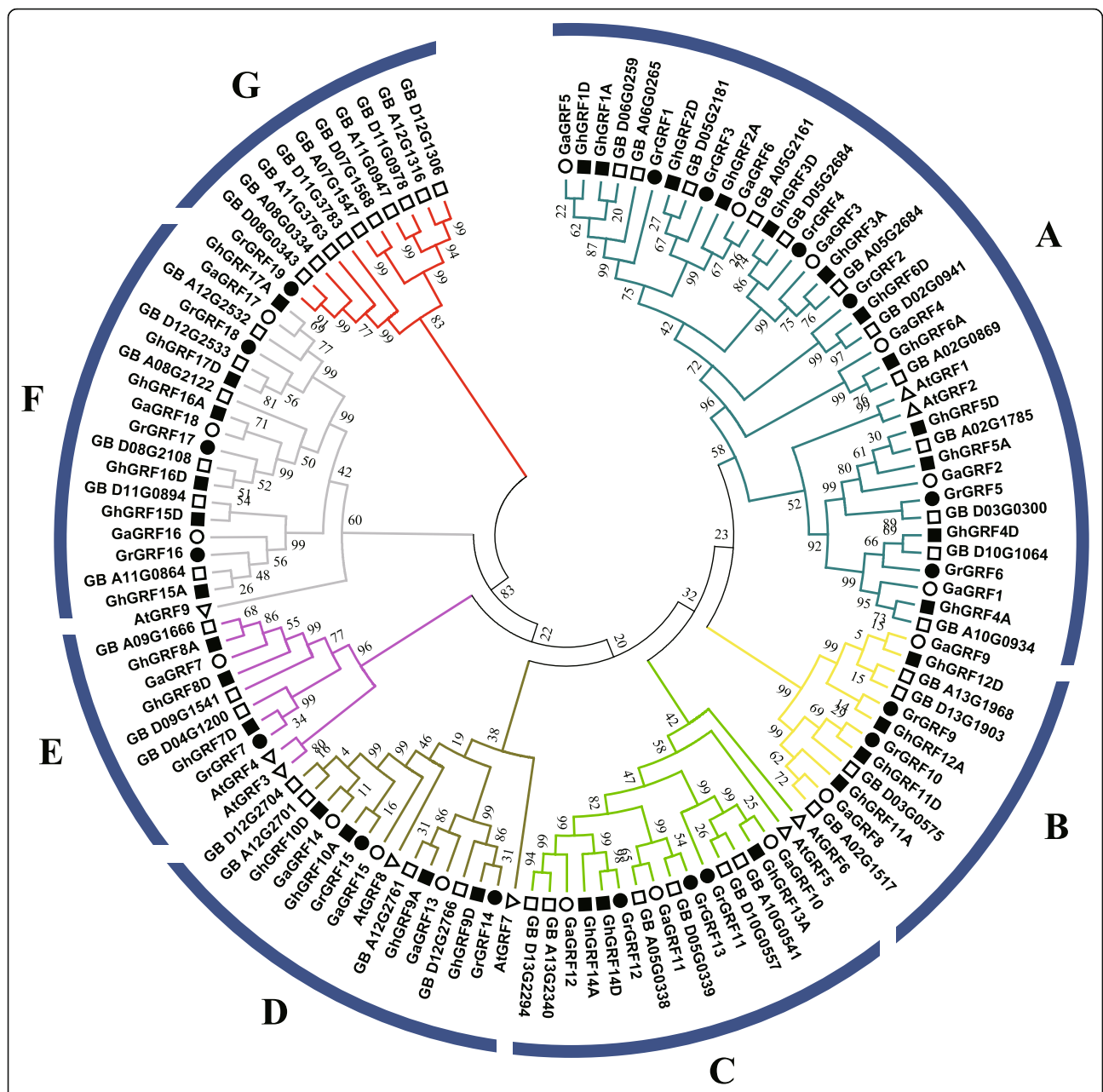Cao *et al. BMC Genomics*   (2020) 21:575

Page 6 of 16

Fig. 4 Phylogenetic analysis of GRF proteins from three *Gossypium* species and *Arabidopsis thaliana*. The phylogenetic tree was established with entire protein sequences with ML methods. The numbers on the branches indicate bootstrap support values from 1000 replications. The scale represents the units of the number of amino acid substitutions per site. The protein sequences used in the phylogenetic analysis are listed in Table 1. **a-g** represent seven phylogenetic clades

Although the GRF family exhibits general preservation, a few deviations were noted. For example, Clade C exhibits evidence of homoeolog loss; that is, the D copy of *GhGRF13* is missing from polyploid genome, as is both copy of *GaGRF11/GrGRF13* (one from each subgenome). In clade D, genes related to *AtGRF8* exhibit duplication in *G. arboreum* only, while the A-homoeolog from either polyploid species was not recovered from the genome sequence. This may indicate a duplication in
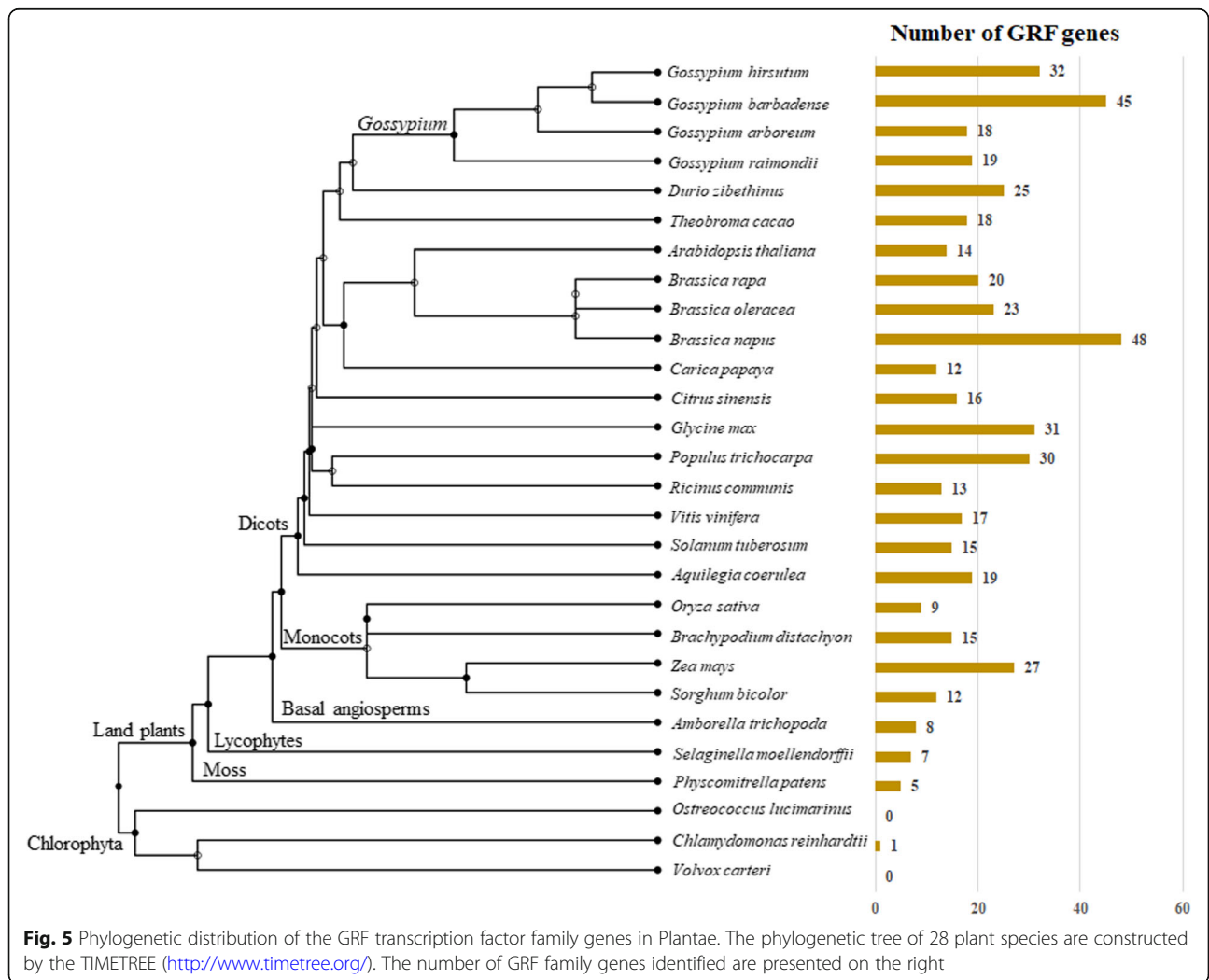
G. arboreum after divergence from the polyploid ancestor coupled with loss in the A-subgenome of the polyploid. The evolution leading to clade E is less straightforward in that it is composed of eight genes, including one from each diploid progenitor and three from each tetraploid species. The phylogenetic topology, however, does not suggest a simple duplication of one homoeologous copy; rather, the homoeologous copies of *GhGRF8* are placed sister first to each other and then to

Cao *et al. BMC Genomics*   (2020) 21:575

Page 7 of 16

G. arboreum only. The remaining GRF from *G. hirsutum*, *GhGRF7D*, is sister to the only *G. raimondii* GRF of this clade and has no inferred homoeologous A_T copy. Given this pattern of relationships and that *G. raimondii* has one additional GRF (relative to *G. arboreum*; see next section), it seems possible that *GhGRF7D* represents a uniquely inherited GRF present in the D ancestor only and that the *G. raimondii* ortholog of *GhGRF8D* was lost after polyploidization. Alternatively, this pattern could reflect a non-syntenic duplication of *GhGRF8D*, followed by conversion of the original *GhGRF8D* into an A-like copy. It bears noting, however, that these observed deviations could be due to errors in the genome sequences; however, the general preservation of GRF copies in the expected relationships suggests general robustness of the analysis. Clade G contains 11 GRF genes, one from *G. raimondii* and the other ten from *G. barbadense*, also indicative of substantial duplication in *G. barbadense*. These new genes belongs to one family and might specially originate in *G. raimondii* and the

tetraploid progenitor of *G. barbadense* or undergone the homoeolog loss in *G. arboreum* and *G. hirsutum*.

Dynamic evolution of *GRF* family genes in plants

We further evaluated the general preservation of GRF genes in plants using 28 representative species and the same search criteria (see methods and Fig. 5). In Chlorophyta, only one *GRF* gene was identified as putatively encoding a GRF protein (in *Chlamydomonas reinhardtii* only); however, all land plant species surveyed recovered a minimum of five putative *GRF* genes in *Physcomitrella patens* and nearly double were found in angiosperms (8 in the basal angiosperm, *Amborella trichopoda*). Among angiosperms, GRF copy number varied from the minimum of 8, in *Amborella trichopoda*, to 48 copies in the tetraploid *Brassica napus*. Four monocot species were included, where between 9 (in rice) to 27 (in maize) copies were detected. Notably, this high copy number in maize is slightly more than double the copy number in sorghum, likely reflective of the duplicated history of

Fig. 5 Phylogenetic distribution of the GRF transcription factor family genes in Plantae. The phylogenetic tree of 28 plant species are constructed by the TIMETREE (http://www.timetree.org/). The number of GRF family genes identified are presented on the right

maize. GRF copy numbers in the eudicot species surveyed varied even more, from 12 putative *GRF* genes in *Carica papaya* to 48 in *Brassica napus*, notably another polyploid species genome.

Comparatively, *Gossypium* GRF copy number is generally stable. We recovered 19 and 18 GRF genes from *G. raimondii* (model D-genome ancestor) and *G. arboreum* (model A-genome ancestor), respectively, dispersed among 10 of the 13 chromosomes (Additional file 4: Figure S2). The allotetraploid cotton species included here both contain a roughly additive number of genes; however, while *G. hirsutum* appears to have lost genes (34 GRF genes, versus the additive expectation of 36), *G. barbadense* has gained 9 additional predicted copies through duplication (45 versus 36; Fig. 5).

Expression changes of *GhGRF* genes in *G. arboreum, G. hirsutum,* and *G. barbadense* under salt stress

GRF genes have been linked to various aspects of growth, development, and response to stress. Long-term effects of salt stress affect all aspects of the plant life cycle, from seed germination through growth and development to future reproductive potential [36]. While the effects of salt stress on roots is often considered [37, 38], equally important are the consequent physiological changes in leaves where the plants exclude sodium ions, potassium/potassium balance, and reduce water loss [39–45]. Previously, a single cotton GRF gene (*GhGRF1* [27];) was associated with responses to salt stress; however, the contributions of the remaining family members to salt stress responses is unknown.

To evaluate the contribution of these *GhGRF* genes to leaf physiology, we first evaluated their expression under normal conditions in 12 different cotton tissues, including root, stem, leaf, cotyledon, hypocotyl, petal, stamen, pistil, sepals, receptacle, ovules (0 dpa) and fiber (6 dpa), using both qPCR and existing RNA-seq data; homoeologs were not distinguished in the qPCR analysis. Most *GhGRF* genes exhibited different expression patterns among the various tissues; however, certain generalities were observed. Several tissues exhibited distinct paralog preference in our qPCR survey. For example, six of the examined *GhGRF* genes (*GhGRF5, 6, 11, 15–17*; Fig. 6) were broadly expressed across tissues, whereas six other *GhGRF* genes (*GhGRF7, 9, 10, 12–14*; Fig. 6) were mainly expressed in the ovule and fiber samples, with limited expression in most other tissues. Most *GhGRF* genes, including *GhGRF5-GhGRF12*, were expressed in the developing ovule (Fig. 6), which transitioned to preferential usage of seven genes (*GhGRF1, GhGRF4, GhGRF7, GhGRF10, GhGRF13* and *GhGRF14*) during the primary elongation stage of cotton fiber development (6 dpa). Transcripts of *GhGRF1, GhGRF2, GhGRF3* and *GhGRF4* were most abundant in stem and root, whereas

GhGRF16 and *GhGRF17* were preferentially expressed in cotyledon. Over half of the *GhGRF* genes showed the relatively high level of transcripts in leaves.

RNA-sequencing (RNA-Seq) data (Additional file 5: Figure S3) obtained from the *G. hirsutum* genome sequencing project [30] and the ccNET database [46] are generally congruent with the expression patterns revealed by qPCR (Fig. 6), although there exist a few discrepancies (Additional file 5: Figure S3). These are generally limited to relative abundance among tissues, such as *GhGRF14*, which exhibited the highest expression in pistil by the RNA-Seq data, but showed higher expression in ovule (0 dpa) and fiber (6 dpa) using qPCR (Fig. 6). Expression in the leaf RNA-seq was much lower for most genes, except *GhGRF1*, and the observation may be attributable to differences in leaf maturity. Generally, however, the two expression analyses suggest a complex pattern of paralog usage among GhGRF genes indicative of multiple biological functions during plant growth and development.

Upland cotton can experience stress [38] due to salt accumulation in the soil. Since *GRF*s have been shown to respond to salt stress responses in other plants [47], we observed phenotypic changes under two concentrations of NaCl and concurrently analyzed *GhGRF* expression profiles using qRT-PCR (Fig. 7). Phenotypic changes in the leaves of Upland cotton were observed using both 200 and 500 mM NaCl over a period of 6 h. With 200 mM NaCl, leaves of *G. hirsutum* experienced yellowing only (indicative of chloroplast degradation), whereas treatment with 500 mM NaCl resulted in leaf yellowing and withering (Fig. 7a). Expression changes (relative to the control; Fig. 7b) were not significant for the 200 mM NaCl at three time-points (1 h, 3 h and 6 h) post-treatment; however, the higher NaCl concentration (500 mM) resulted in down-regulation of five genes (*GhGRF3, GhGRF4, GhGRF5, GhGRF7,* and *GhGRF16*; Fig. 7b). No *GhGRF* genes experienced up-regulation under salt stress, which is congruent with the observation that salt stress leads to diminished growth. The RNA-seq data derived from the *G. hirsutum* genome and ccNET [46] generally concur with these results. For example, 3 of the 5 genes downregulated here (i.e., *GhGRF3, GhGRF4,* and *GhGRF16*) also exhibited significantly decreased expression levels (≥2-fold change) at four time-points during NaCl stress in the available RNA-seq data (Fig. 8). These three GRFs in particular may be commonly responsive to salt stress and warrant further functional characterization.

We also observed the phenotypic changes of *G. arboreum* and *G. barbadense* under two concentrations of NaCl and concurrently analyzed the *GaGRF* and *GbGRF* expression profiles by qRT-PCR (Fig. 9). Phenotypic changes in the leaves of *G. arboreum* and *G. barbadense* were observed using both 200 and 500 mM NaCl over a
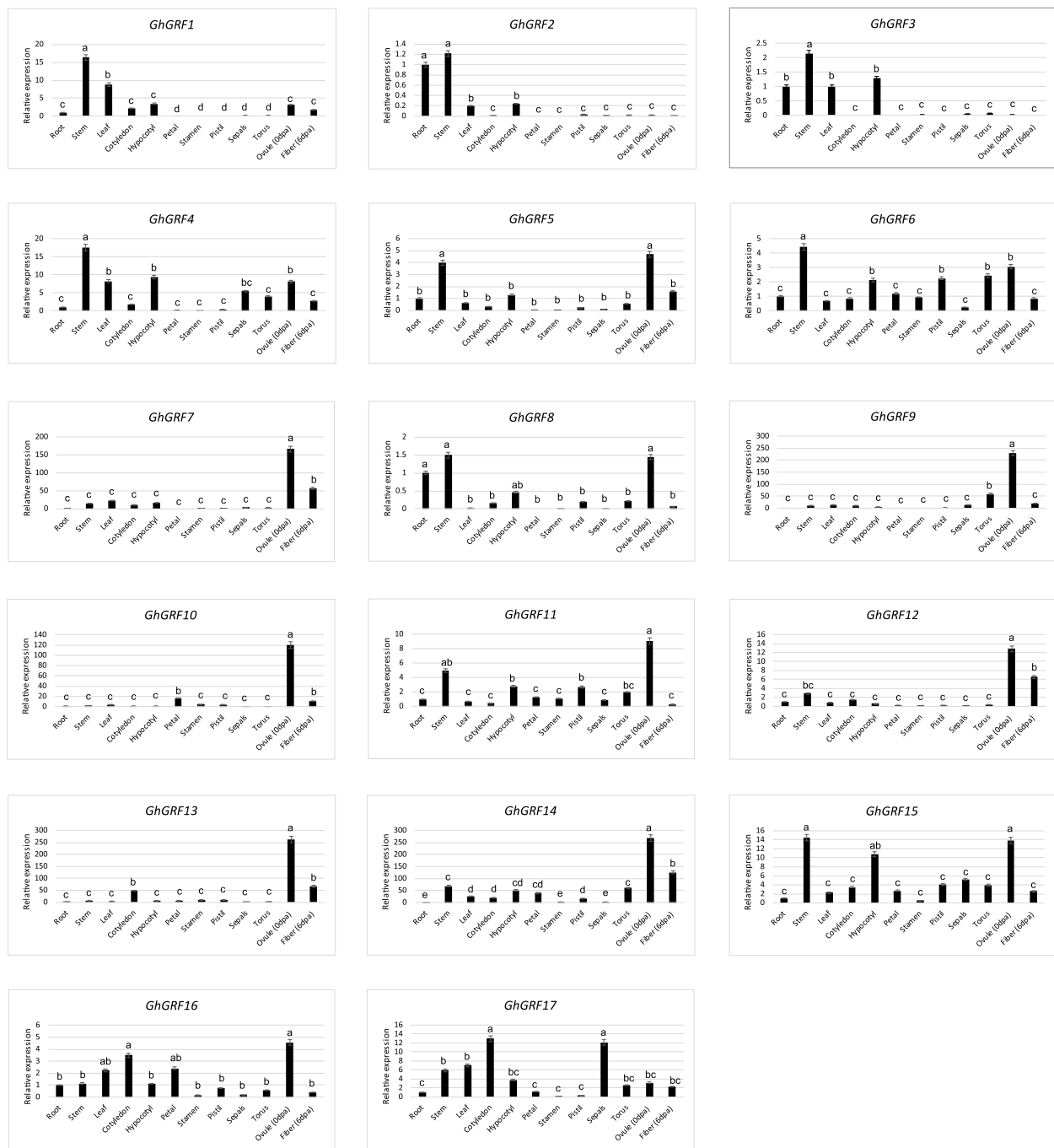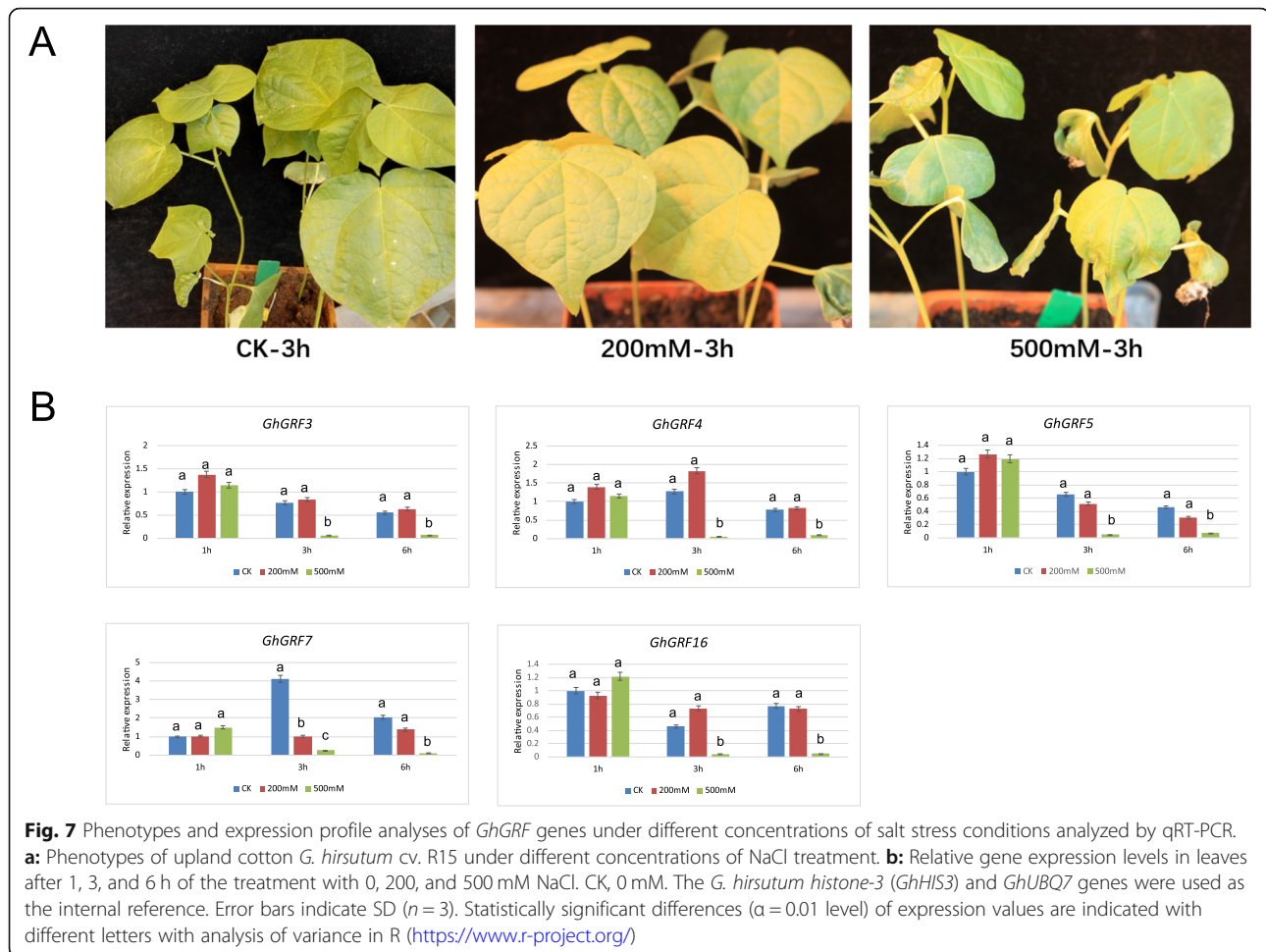
Fig. 6 Expression of *GhGRF* genes in different tissues analyzed by real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR). Twelve tissues, including roots, stems, leaves, cotyledon, hypocotyl petal, stamen, pistil, sepals, torus, ovules (0dpa) and fiber (6dpa), were collected from the continuing growing cotton plants. Relative gene expression levels are normalized to *histone-3* gene (*GhHIS3*) values. Error bars indicate SD (*n* = 3). Statistically significant differences ("a" is different from "b" or "c", α = 0.01 level) of expression values are indicated with different letters with analysis of variance in R (https://www.r-project.org/)

period of 6 h. Although neither 200 nor 500 mM NaCl treatment resulting in yellowing or withering of *G. arboreum* leaves (Fig. 9a), downregulation of six *GaGRF* genes (Fig. 9b) was nevertheless detected at three time-points (1 h, 3 h and 6 h) post-treatment. Notably, only

two of these genes, i.e., *GaGRF5* and *GaGRF16*, were also down-regulated in *G. hirsutum*. Similar to *G. hirsutum*, *G. barbadense* was modestly affected by low salt treatment, with leaf withering beginning at 200 mM NaCl and both leaf yellowing and withering present at

Fig. 7 Phenotypes and expression profile analyses of *GhGRF* genes under different concentrations of salt stress conditions analyzed by qRT-PCR. **a:** Phenotypes of upland cotton *G. hirsutum* cv. R15 under different concentrations of NaCl treatment. **b:** Relative gene expression levels in leaves after 1, 3, and 6 h of the treatment with 0, 200, and 500 mM NaCl. CK, 0 mM. The *G. hirsutum histone-3* (*GhHIS3*) and *GhUBQ7* genes were used as the internal reference. Error bars indicate SD (*n* = 3). Statistically significant differences (α = 0.01 level) of expression values are indicated with different letters with analysis of variance in R (https://www.r-project.org/)
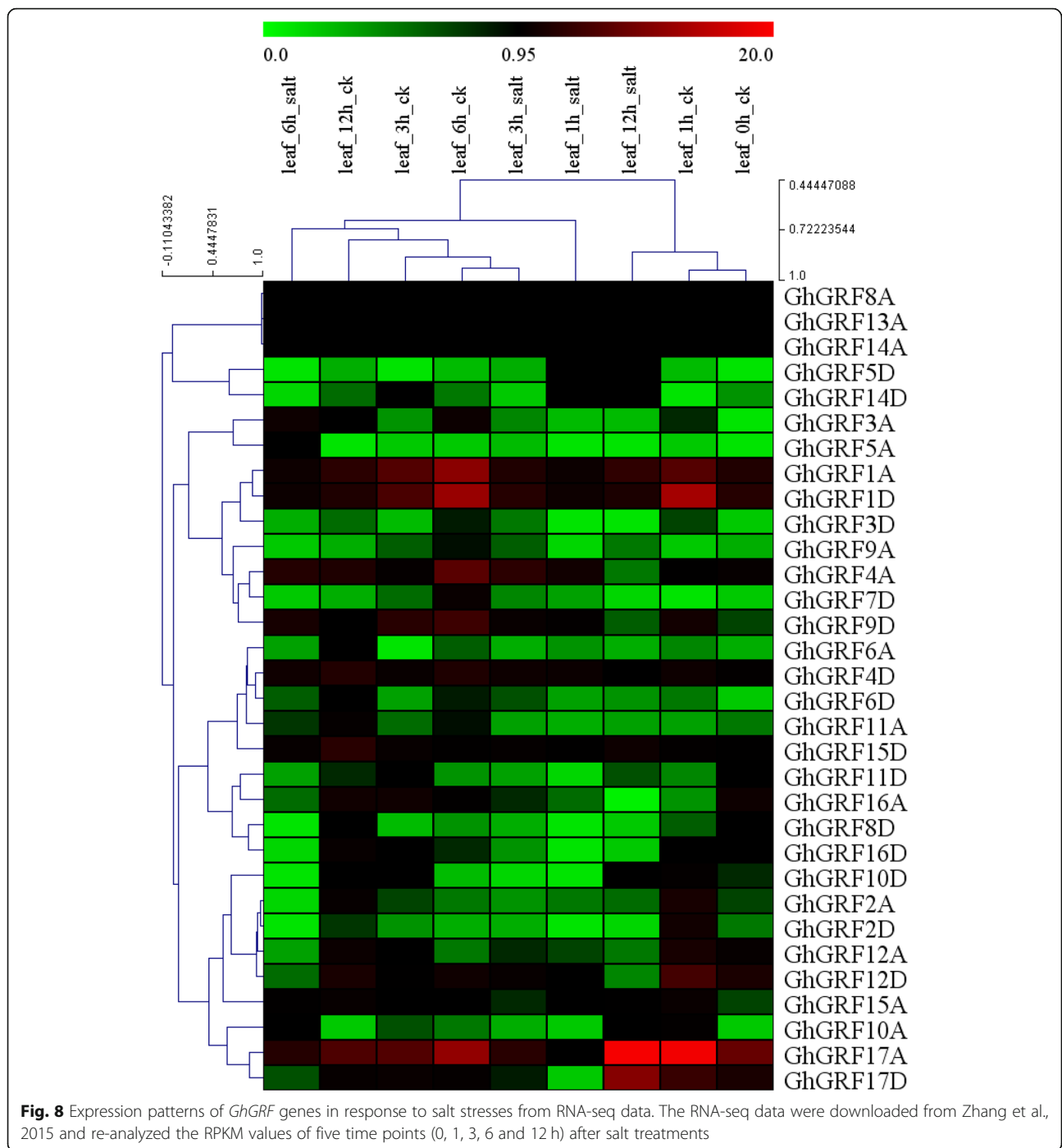
500 mM NaCl (Fig. 9c); however, here only three *GbGRF* genes (Fig. 9d) exhibited down-regulation, only one of which (*GbGRF4*) exhibited downregulation in either *G. hirsutum* or *G. arboreum*. Of the three genes downregulated in *G. barbadense*, expression changes of *GbGRF17* was significant for the 200 mM and 500 mM NaCl at three time-points (1 h, 3 h and 6 h) post-treatment. These results demonstrate that (1) *G. arboreum* has better salt tolerance than both *G. hirsutum* and *G. barbadense*, congruent with the previous observations [28, 48, 49], and (2) GRF responses to salt stress overlap, but are not consistent, among species.

Discussion

The GRF transcription factor family of genes has been investigated in a number of plant species, including *Arabidopsis thaliana* [20, 50–53], *Zea mays* [35], *Brassica napus* [54], *Brassica rapa* [55], *Oryza sativa* [56–58], *Brachypodium distachyon* [59] and *Solanum lycopersicum* [22]. The family generally contains conserved functional regions, including a protein-protein interaction QLQ (glutamine, leucine, glutamine) domain that was involved in chromatin remodeling [60] and a plant-specific WRC (tryptophan, arginine, cysteine) motif relevant to DNA binding and targeting of the TF to the nucleus [14, 58]. The conservation of these domains among GRF genes facilitates computational identification of the family from new and emerging genomes. Here, we identify 32 GRF genes from the allotetraploid *G. hirsutum*, and show that these genes are generally conserved among the diploid progenitors and the cultivated allotetraploid cottons. Analysis of the GRF genes in an additional cotton species, as well as representatives of diverse plant lineages, show that the family is somewhat labile but the most remarkable changes in copy number are attributable to retention after genome duplication. In tetraploid cottons, the gene copy number is nearly additive in *G. hirsutum,* but contains 25% more paralogs in the related (and also domesticated) *G. barbadense* genome. Phylogenetic analysis indicated that these additive genes belongs to one family and specially originate in *G. barbadense.*
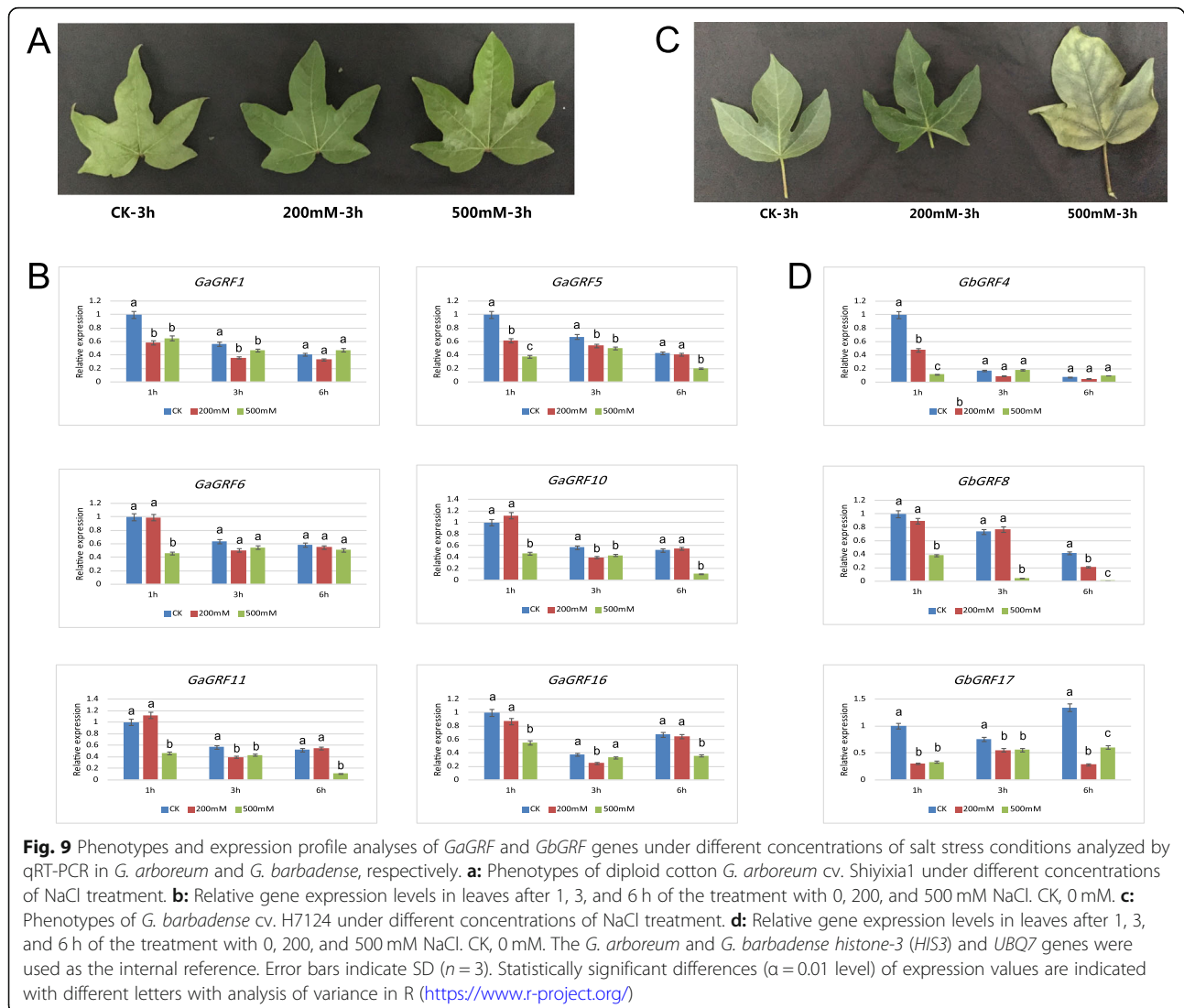
Cao et al. BMC Genomics (2020) 21:575

Page 11 of 16

Fig. 8 Expression patterns of *GhGRF* genes in response to salt stresses from RNA-seq data. The RNA-seq data were downloaded from Zhang et al., 2015 and re-analyzed the RPKM values of five time points (0, 1, 3, 6 and 12 h) after salt treatments

Further research is required to discern what functional relevance these additional copies have.

Conclusions

We systematically characterized cotton GRF family genes using bioinformatic and phylogenetic approaches and gene expression analyses. We analyzed gene structures, chromosomal locations, intron-exon organizations, phylogenetic relationships and expression profile patterns in different cotton tissues and under salt stress condition to predict their possible biological functions. *GhGRF* genes are variably expressed in different cotton tissues with particularly high expression in ovules. The decreased expression of several *GhGRF* genes in response to salt stress treatments implies their function in the regulations of growth and development under the abiotic stress conditions. Together, our results provide data to facilitate the functional identification of the GRF

Fig. 9 Phenotypes and expression profile analyses of *GaGRF* and *GbGRF* genes under different concentrations of salt stress conditions analyzed by qRT-PCR in *G. arboreum* and *G. barbadense*, respectively. **a:** Phenotypes of diploid cotton *G. arboreum* cv. Shiyixia1 under different concentrations of NaCl treatment. **b:** Relative gene expression levels in leaves after 1, 3, and 6 h of the treatment with 0, 200, and 500 mM NaCl. CK, 0 mM. **c:** Phenotypes of *G. barbadense* cv. H7124 under different concentrations of NaCl treatment. **d:** Relative gene expression levels in leaves after 1, 3, and 6 h of the treatment with 0, 200, and 500 mM NaCl. CK, 0 mM. The *G. arboreum* and *G. barbadense histone-3* (*HIS3*) and *UBQ7* genes were used as the internal reference. Error bars indicate SD ($n = 3$). Statistically significant differences ($α = 0.01$ level) of expression values are indicated with different letters with analysis of variance in R (https://www.r-project.org/)

genes in cotton plant growth, development and stress tolerance.

Methods

Identification of GRF family genes and GRF proteins in diploid and tetraploid *Gossypium* species

We downloaded the genome sequences of four cotton species from the CottonGen database [61], including *G. raimondii* [25], *G. arboreum* [27], *G. hirsutum* [30], and *G. barbadense* [62]. To identify all putative GRF transcription factor proteins in each genome assembly, the GRF protein conserved domains (PF08879 for WRC domain and PF08880 for QLQ domain) were used to develop a Hidden Markov Model [63] profile matrix via the hmmbuild program from the HMMER package [64] using default parameters. This HMM profile matrix was used in conjunction with hmmsearch with default parameters against three *Gossypium* genome databases,

i.e., *G. raimondii*, *G. arboreum,* and *G. hirsutum*, to identify putative GRF genes (*GhGRF*s) that contain the relevant conserved protein domains. Genes were considered candidate GRFs if they harbored WRC and QLQ domains within the N-terminal [13]. Previously identified *GRF* gene sequences from *Arabidopsis thaliana* (*AtGRF*s) were retrieved from the TAIR database [64] for phylogenetic comparison. The presence of conserved domains in each *Arabidopsis* gene was verified using the SMART conserved domain search tool [65] and Pfam databases [66]. The same method was used to identify the number of GRF genes in other plant genera/species (Additional file 1: Table S1).

Chromosomal location and gene structure analyses

Chromosomal locations for each of the above identified *GhGRFs* were extracted from the genome annotation gff3 file [30]. Chromosomal locations of the predicted

*GhGRF*s was visualized using MapChart [67], and the exon-intron structure of each gene was displayed using the online tool GSDS 2.0 [68]. The number of amino acids, molecular weight (MW), and theoretical isoelectric point (pI) of putative *GhGRF* proteins were determined using the ProtParam tool [69].

Sequence alignment and phylogenetic analyses

Complete protein-coding sequences for each of these genes from all three cotton species and *Arabidopsis* were aligned using MAFFT with the G-INS-i algorithm [70]. Phylogenetic analyses based on the whole protein sequences were performed using Neighbor-Joining (NJ) and Maximum Likelihood (ML). The NJ tree was constructed using MEGA version 5.03 [71] by sampling 1000 bootstrap replicates. The ML tree was also built using MEGA version 5.03, using the general time reversible (GTR) model, including rate variation among sites (+G) and invariable sites (+I; full model = GTR + G + I), and running 1000 bootstrap replicates of the data.

Transcriptome data-based gene expression analyses

Raw RNA-Seq data for *G. hirsutum* seed, root, stem, leaf, torus, petal, stamen, ovary, calyx, ovule (− 3 dpa, − 1 dpa, 0 dpa, 1 dpa, 3 dpa, 5 dpa, 10 dpa, 20 dpa, 25dpa, 35dpa) and fiber (5 dpa, 10 dpa, 20 dpa, 25dpa) were downloaded from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA248163) [30], represented by one library each. Reads were mapped to *G. hirsutum* genome [30] via HISAT2 software with default parameters, and read abundance with calculated via StringTie [72, 73]. Read counts were normalized in R3.2 using RUVSeq [74] and the internal control reference gene *GhUBQ7* (accession number: DQ116441), which is detected at relatively constant levels across different cotton samples [75]. Potential batch-effects were corrected by an improved version of ComBat, ComBat-seq [76]. Gene expression was estimated by Ballgown [86], using fragments per kilobase million (FPKM) values to calculate the gene expression levels across libraries. Expression levels of *G. hirsutum* leaf RNA-Seq data (in FPKM) for each *GhGRF* gene under salt stress (time points: 0, 1, 3, 6, 12 h) were retrieved from the ccNET database [46]. Genes were considered differentially expressed if expression varied more than two-fold change with a *p*-value of less than 0.05. Multiple Experiment Viewer (MeV) [77] was used to display the gene expression patterns from the reported FPKM values.

Plant cultivation and treatment

To generate new expression information via qRT-PCR, we grew representatives of *G. hirsutum, G. arboreum,* and *G. barbadense.* For *G. hirsutum,* seeds of *G. hirsutum* cv. R15

[78], were germinated in potting soil in a growth chamber, and the resulting seedlings were maintained in a controlled environment at 28 °C day/20 °C night, with a 16-h light/8-h dark photoperiod. Roots, stems, leaves, cotyledons, and hypocotyls were collected from the three-week old plants, and additional samples were collected from older, flowering plants; these include petal, stamen, pistil, sepals, torus, ovules (0 dpa (days post anthesis)) and fiber (6 dpa). Three biological replicates were collected for each sample, each with three technical replicates. For salt treatment, 28-day old plants were sprayed with 200 and 500 mM NaCl solution after surfactant (Triton X-100) treatments. Leaves from salt-treated plants were collected at 0 (control), 1, 3, and 6 h post-NaCl treatment for further expression analyses. All plant tissues were frozen in liquid nitrogen immediately after collection and stored at − 80 °C until RNA extraction. All treatments was sampled at least three times.

Similarly, *G. arboreum* cv. Shixiya1 and *G. barbadense* H7124 were grown for qRT-PCR of salt-exposed leaf tissue timepoints only. For this experiment, seeds of *G. arboreum* cv. Shixiya1 were provided by Prof. Tianzhen Zhang and *G. barbadense* H7124 seeds were provided from the Esquel Group. These two *Gossypium* species were planted in the Damao field in Sanya, Hainan Province, China. For salt treatment, 50-day old plants were sprayed with 200 and 500 mM NaCl solution after surfactant (Triton X-100) treatment. Leaves from salt-treated plants were collected at 0 (control), 1, 3, and 6 h post-NaCl treatment as above.

RNA extraction, cDNA synthesis and qRT-PCR expression analyses

Total RNAs from cotton tissues were extracted using the RNAprep pure plant kit (TIANGEN, Shanghai, China) according to the manufacturer's protocol. The resulting RNAs were treated with DNase I prior to synthesizing cDNA with oligo (dT) primers and M-MLV Reverse Transcriptase (Invitrogen); these products were diluted 5-fold before use. For quantitative real-time PCR (qRT-PCR), Primer5 software was used to design gene-specific forward and reverse primers (Additional file 2: Table S2). As these primers are not homoeolog specific, both copies were amplified when retained in duplicate. Analyses were performed with SYBR-Green PCR Mastermix (TaKaRa) on a cycler (Mastercycler RealPlex; Eppendorf Ltd., Shanghai, China). The *G. hirsutum* histone-3 (*GhHIS3*, AF024716) and *GhUBQ7* (accession number: DQ116441) genes were used as internal references, and the relative amount of amplified product was calculated following the 2-ΔΔCt method [79]. For the *G. hirsutum* samples, relative expression levels among different organs were normalized by calibrating with the root sample from that plant. The root sample was washed with DEPC sterile water three times before extracting the RNA.

Cao *et al. BMC Genomics*   (2020) 21:575

Page 14 of 16

Supplementary information

Additional file 1: Table S1. 28 plant species used for identifying the number of GRF family genes.

Additional file 2: Table S2. List of forward and reverse primers used for qRT-PCR analyses

Additional file 3: Figure S1. Multiple alignment of GhGRF (*Gossypium hirsutum* GRF) protein sequences

Additional file 4: Figure S2. Chromosome distribution of GaGRF (*Gossypium arboreum* GRF) and GrGRF (*Gossypium raimondii* GRF) genes.

Additional file 5: Figure S3. Expression patterns of *GRF* genes in *G. hirsutum* using RNA-seq data. The RNA-seq data expression profiles were from Zhang et al., (2015) and ccNET database [46]. FPKM represents fragments per kilobase of exon model per million mapped reads. DPA, days post anthesis

Abbreviations
GRFs: Growth regulating factors; DPA: Days post anthesis; FPKM: Fragments per kilobase of transcript per million mapped fragments; *G. arboreum*: *Gossypium arboreum*; *G. hirsutum*: *Gossypium hirsutum*; *G. raimondii*: *Gossypium raimondii*; MW: Molecular weight; pI: Isoelectric point; qRT-PCR: Quantitative real-time polymerase chain reaction

Authors' contributions
ZWC and CEG conceived the research. JFC, CCH, XL, ZSZ, YGZ and ZWC performed the experiments. JQH, XFZ, XXSG, LJW, CEG and ZWC contributed materials and analyzed data. ZWC, JFC and CEG wrote the manuscript. JFW participated in discussion and revised the manuscript. All authors have read and approved the manuscript.

Availability of data and materials
The genome sequences of four cotton species and the genome annotation gff3 file were downloaded from the CottonGen database (https://www.cottongen.org/data/download) [61]. GRF gene sequences from *Arabidopsis thaliana* were retrieved from the TAIR database (https://www.arabidopsis.org/servlets/Search?type=general&search_action=detail&method=1&show_obsolete=F&name=GRF&sub_type=gene&SEARCH_EXACT=4&SEARCH_CONTAINS=1) [65]. Raw RNA-Seq data for *G. hirsutum* seed, root, stem, leaf, torus, petal, stamen, ovary, calyx, ovule and fiber were downloaded from the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA248163) (NCBI Sequence Read Archive SRR1695173, SRR1695174, SRR1695175, SRR1695177, SRR1695178, SRR1695179, SRR1695181, SRR1695182, SRR1695183, SRR1695184, SRR1695185, SRR1695191, SRR1695192, SRR1695193,SRR1695194, SRR1768504, SRR1768505, SRR1768506, SRR1768507, SRR1768508, SRR1768509, SRR1768510, SRR1768511, SRR1768512, SRR1768513, SRR1768514, SRR1768515, SRR1768516, SRR1768517, SRR1768518 and SRR1768519) [30]. The *G. hirsutum* histone-3 (GhHIS3, AF024716) and GhUBQ7 (accession number: DQ116441) genes were downloaded from the National Center for Biotechnology Information (NCBI) database, which were used as internal references. All other data generated or analyzed during this study are included in this published 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
¹National Key Laboratory of Plant Molecular Genetics and National Center for Plant Gene Research, Institute of Plant Physiology and Ecology/CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China. ²Plant Stress Biology Center, Institute of Plant Physiology and Ecology/CAS Center for Excellence in Molecular Plant Sciences, Chinese Academy of Sciences, Shanghai 200032, China. ³University of Chinese Academy of Sciences, Shanghai 200032, China. ⁴Esquel Group, 25 Harbour Road, Wanchai, Hong Kong, China. ⁵School of Life Science and Technology, ShanghaiTech University, Shanghai 201210, China. ⁶Department of Ecology, Evolution and Organismal Biology, Iowa State University, Ames, IA 50011, USA. ⁷Institute of Carbon Materials Science, Shanxi Datong University, Datong 037009, China.

References
1. Wendel FJ, Grover C. Taxonomy and evolution of the cotton genus, Gossypium. 2nd ed. Madison: American Society of Agronomy Inc., Crop Science Society of America Inc., and Soil Science Society of America Inc; 2015.
2. Wang K, Wendel JF, Hua J. Designations for individual genomes and chromosomes in Gossypium. J Cotton Res. 2018;1(1):3.
3. Tian X, Ruan JX, Huang JQ, Yang CQ, Fang X, Chen ZW, Hong H, Wang LJ, Mao YB, Lu S, et al. Characterization of gossypol biosynthetic pathway. Proc Natl Acad Sci U S A. 2018;115(23):E5410–8.
4. van Kretschmar B, Cabrera AR, Bradley JR, Roe RM. Novel adult feeding disruption test (FDT) to detect insecticide resistance of lepidopteran pests in cotton. Pest Manag Sci. 2013;69(5):652–60.
5. Chen D, Chen F, Chen C, Chen X, Mao Y. Transcriptome analysis of three cotton pests reveals features of gene expressions in the mesophyll feeder Apolygus lucorum. Sci China Life Sci. 2017;60(8):826–38.
6. Zhang W, Lu Y, van der Werf W, Huang J, Wu F, Zhou K, Deng X, Jiang Y, Wu K, Rosegrant MW. Multidecadal, county-level analysis of the effects of land use, Bt cotton, and weather on cotton pests in China. Proc Natl Acad Sci U S A. 2018;115(33):E7700–9.
7. Huang JQ, Fang X, Tian X, Chen P, Lin JL, Guo XX, Li JX, Fan Z, Song WM, Chen FY, et al. Aromatization of natural products by a specialized detoxification enzyme. Nat Chem Biol. 2020.
8. Deinlein U, Stephan AB, Horie T, Luo W, Xu GH, Schroeder JI. Plant salt-tolerance mechanisms. Trends Plant Sci. 2014;19(6):371–9.
9. Liang W, Ma X, Wan P, Liu L. Plant salt-tolerance mechanism: a review. Biochem Biophys Res Commun. 2018;495(1):286–91.
10. Parihar P, Singh S, Singh R, Singh VP, Prasad SM. Effect of salinity stress on plants and its tolerance strategies: a review. Environ Sci Pollut Res Int. 2015; 22(6):4056–75.
11. Fahad S, Hussain S, Matloob A, Khan FA, Khaliq A, Saud S, Hassan S, Shan D, Khan F, Ullah N, et al. Phytohormones and plant responses to salinity stress: a review. Plant Growth Regul. 2015;75(2):391–404.
12. Mickelbart MV, Hasegawa PM, Bailey-Serres J. Genetic mechanisms of abiotic stress tolerance that translate to crop yield stability. Nat Rev Genet. 2015; 16(4):237–51.
13. Omidbakhshfard MA, Proost S, Fujikura U, Mueller-Roeber B. Growth-regulating factors (GRFs): a small transcription factor family with important functions in plant biology. Mol Plant. 2015;8(7):998–1010.
14. van der Knaap E, Kim JH, Kende H. A novel gibberellin-induced gene from rice and its potential regulatory role in stem growth. Plant Physiol. 2000; 122(3):695–704.
15. Bao ML, Bian HW, Zha YL, Li FY, Sun YZ, Bai B, Chen ZH, Wang JH, Zhu MY, Han N. miR396a-mediated basic helix-loop-helix transcription factor bHLH74

Cao *et al. BMC Genomics*   (2020) 21:575

Page 15 of 16

repression acts as a regulator for root growth in Arabidopsis seedlings. Plant Cell Physiol. 2014;55(7):1343–53.

16. Liang G, He H, Li Y, Wang F, Yu DQ. Molecular mechanism of microRNA396 mediating pistil development in Arabidopsis. Plant Physiol. 2014;164(1):249–58.

17. Liu HH, Guo SY, Xu YY, Li CH, Zhang ZY, Zhang DJ, Xu SJ, Zhang C, Chong K. OsmiR396d-regulated OsGRFs function in floral organogenesis in Rice through binding to their targets OsJMJ706 and OsCR4. Plant Physiol. 2014;165(1):160–74.

18. Debernardi JM, Mecchia MA, Vercruyssen L, Smaczniak C, Kaufmann K, Inze D, Rodriguez RE, Palatnik JF. Post-transcriptional control of GRF transcription factors by microRNA miR396 and GIF co-activator affects leaf size and longevity. Plant J. 2014;79(3):413–26.

19. Hewezi T, Maier TR, Nettleton D, Baum TJ. The Arabidopsis MicroRNA396-GRF1/GRF3 regulatory module acts as a developmental regulator in the reprogramming of root cells during cyst nematode infection. Plant Physiol. 2012;159(1):321–35.

20. Kim JS, Mizoi J, Kidokoro S, Maruyama K, Nakajima J, Nakashima K, Mitsuda N, Takiguchi Y, Ohme-Takagi M, Kondou Y, et al. Arabidopsis growth-regulating Factor7 functions as a transcriptional repressor of abscisic acid-and osmotic stress-responsive genes, including DREB2A. Plant Cell. 2012;24(8):3393–405.

21. Casadevall R, Rodriguez RE, Debernardi JM, Palatnik JF, Casati P. Repression of growth regulating factors by the microRNA396 inhibits cell proliferation by UV-B radiation in Arabidopsis leaves. Plant Cell. 2013;25(9):3570–83.

22. Khatun K, Robin AHK, Park JI, Nath UK, Kim CK, Lim KB, Nou IS, Chung MY. Molecular characterization and expression profiling of tomato GRF transcription factor family genes in response to abiotic stresses and phytohormones. Int J Mol Sci. 2017;18(5):1056.

23. Heidel AJ, Clarke JD, Antonovics J, Dong XN. Fitness costs of mutations affecting the systemic acquired resistance pathway in *Arabidopsis thaliana*. Genetics. 2004;168(4):2197–206.

24. Sakuma Y, Maruyama K, Osakabe Y, Qin F, Seki M, Shinozaki K, Yamaguchi-Shinozaki K. Functional analysis of an Arabidopsis transcription factor, DREB2A, involved in drought-responsive gene expression. Plant Cell. 2006;18(5):1292–309.

25. 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.

26. Wang K, Wang Z, Li F, Ye W, Wang J, Song G, Yue Z, Cong L, Shang H, Zhu S, et al. The draft genome of a diploid cotton Gossypium raimondii. Nat Genet. 2012;44(10):1098–103.

27. Li F, Fan G, Wang K, Sun F, Yuan Y, Song G, Li Q, Ma Z, Lu C, Zou C, et al. Genome sequence of the cultivated cotton Gossypium arboreum. Nat Genet. 2014;46(6):567–72.

28. 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.

29. Li FG, Fan GY, Lu CR, Xiao GH, Zou CS, Kohel RJ, Ma ZY, Shang HH, Ma XF, Wu JY, et al. Genome sequence of cultivated upland cotton (Gossypium hirsutum TM-1) provides insights into genome evolution. Nat Biotechnol. 2015;33(5):524–U242.

30. Zhang T, Hu Y, Jiang W, Fang L, Guan X, Chen J, Zhang J, Saski CA, Scheffler BE, Stelly DM, et al. Sequencing of allotetraploid cotton (Gossypium hirsutum L. acc. TM-1) provides a resource for fiber improvement. Nat Biotechnol. 2015;33(5):531–7.

31. Wang M, Tu L, Yuan D, Zhu SC, Li J, Liu F, Pei L, Wang P, Zhao G, et al. Reference genome sequences of two cultivated allotetraploid cottons, Gossypium hirsutum and Gossypium barbadense. Nat Genet. 2019;51(2):224–9.

32. Liu X, Zhao B, Zheng HJ, Hu Y, Lu G, Yang CQ, Chen JD, Chen JJ, Chen DY, Zhang L, et al. *Gossypium barbadense* genome sequence provides insight into the evolution of extra-long staple fiber and specialized metabolites. Sci Rep-Uk. 2015;5:14139.

33. Yuan D, Tang Z, Wang M, Gao W, Tu L, Jin X, Chen L, He Y, Zhang L, Zhu L, et al. The genome sequence of Sea-Island cotton (Gossypium barbadense) provides insights into the allopolyploidization and development of superior spinnable fibres. Sci Rep. 2015;5:17662.

34. 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.

35. Zhang DF, Li B, Jia GQ, Zhang TF, Dai JR, Li JS, Wang SC. Isolation and characterization of genes encoding GRF transcription factors and GIF transcriptional coactivators in maize (*Zea mays* L.). Plant Sci. 2008;175(6):809–17.

36. Dong Y, Hu G, Yu J, Thu SW, Grover CE, Zhu S, Wendel JF. Salt-tolerance diversity in diploid and polyploid cotton (*Gossypium*) species. Plant J. 2020;101(5):1135–51.

37. Shi GY, Guo XY, Guo JY, Liu LH, Hua JP. Analyzing serial cDNA libraries revealed reactive oxygen species and gibberellins signaling pathways in the salt response of upland cotton (*Gossypium hirsutum* L.). Plant Cell Rep. 2015;34(6):1005–23.

38. Guo JY, Shi GY, Guo XY, Zhang LW, Xu WY, Wang YM, Su Z, Hua JP. Transcriptome analysis reveals that distinct metabolic pathways operate in salt-tolerant and salt-sensitive upland cotton varieties subjected to salinity stress. Plant Sci. 2015;238:33–45.

39. Mitsuya S, Kawasaki M, Taniguchi M, Miyake H. Relationship between salinity-induced damages and aging in rice leaf tissues. Plant Prod Sci. 2003;6(3):213–8.

40. Chazen O, Hartung W, Neumann PM. The different effects of Peg-6000 and Nacl on leaf development are associated with differential inhibition of root water transport. Plant Cell Environ. 1995;18(7):727–35.

41. Shabala S, Demidchik V, Shabala L, Cuin TA, Smith SJ, Miller AJ, Davies JM, Newman IA. Extracellular Ca2+ ameliorates NaCl-induced K+ loss from Arabidopsis root and leaf cells by controlling plasma membrane K+-permeable channels. Plant Physiol. 2006;141(4):1653–65.

42. Speer M, Kaiser WM. Ion relations of Symplastic and Apoplastic space in leaves from Spinacia-Oleracea L and Pisum-Sativum L under salinity. Plant Physiol. 1991;97(3):990–7.

43. Wu HH, Shabala L, Barry K, Zhou MX, Shabala S. Ability of leaf mesophyll to retain potassium correlates with salinity tolerance in wheat and barley. Physiol Plantarum. 2013;149(4):515–27.

44. Wu HH, Shabala L, Zhou MX, Shabala S. Chloroplast-generated ROS dominate NaCl- induced K+ efflux in wheat leaf mesophyll. Plant Signal Behav. 2015;10(5).

45. Wu HHY, Zhu M, Shabala L, Zhou MX, Shabala S. K+ retention in leaf mesophyll, an overlooked component of salinity tolerance mechanism: a case study for barley. J Integr Plant Biol. 2015;57(2):171–85.

46. You Q, Xu WY, Zhang K, Zhang LW, Yi X, Yao DX, Wang CC, Zhang XY, Zhao XH, Provart NJ, et al. ccNET: database of co-expression networks with functional modules for diploid and polyploid *Gossypium*. Nucleic Acids Res. 2017;45(9):5625–6.

47. Chen L, Luan Y, Zhai J. Sp-miR396a-5p acts as a stress-responsive genes regulator by conferring tolerance to abiotic stresses and susceptibility to *Phytophthora nicotianae* infection in transgenic tobacco. Plant Cell Rep. 2015;34(12):2013–25.

48. Dilnur T, Peng Z, Pan Z, Palanga KK, Jia Y, Gong W, Du X. Association analysis of salt tolerance in Asiatic cotton (*Gossypium arboreum*) with SNP markers. Int J Mol Sci. 2019;20(9):2168.

49. Liu F, Zhou ZL, Wang CY, Wang YH, Cai XY, Wang XX, Zhang ZS, Wang KB. Genetic diversity and relationship analysis of *Gossypium arboreum* accessions. Genet Mol Res. 2015;14(4):14522–9.

50. Yang FX, Liang G, Liu DM, Yu DQ. Arabidopsis MiR396 mediates the development of leaves and flowers in transgenic tobacco. J Plant Biol. 2009;52(5):475–81.

51. Kim JH, Choi DS, Kende H. The AtGRF family of putative transcription factors is involved in leaf and cotyledon growth in Arabidopsis. Plant J. 2003;36(1):94–104.

52. Kim JH, Lee BH. Growth-regulating Factor4 of *Arabidopsis thaliana* is required for development of leaves, cotyledons, and shoot apical meristem. J Plant Biol. 2006;49(6):463–8.

53. Lee BH, Ko JH, Lee S, Lee Y, Pak JH, Kim JH. The Arabidopsis GRF-interacting factor gene family performs an overlapping function in determining organ size as well as multiple developmental properties. Plant Physiol. 2009;151(2):655–68.

54. Liu J, Hua W, Yang HL, Zhan GM, Li RJ, Deng LB, Wang XF, Liu GH, Wang HZ. The *BnGRF2* gene (*GRF2-like* gene from *Brassica napus*) enhances seed oil production through regulating cell number and plant photosynthesis. J Exp Bot. 2012;63(10):3727–40.

55. Wang FD, Qiu NW, Ding Q, Li JJ, Zhang YH, Li HY, Gao JW. Genome-wide identification and analysis of the growth-regulating factor family in Chinese cabbage (*Brassica rapa* L. ssp pekinensis). BMC Genomics. 2014;15:807.

Cao *et al. BMC Genomics*   (2020) 21:575

Page 16 of 16

56. Kuijt SJH, Greco R, Agalou A, Shao JX, CCJ t H, Overnas E, Osnato M, Curiale S, Meynard D, van Gulik R, et al. Interaction between the growth-regulating factor and Knotted1-like homeobox families of transcription factors1. Plant Physiol. 2014;164(4):1952–66.

57. Luo AD, Liu L, Tang ZS, Bai XQ, Cao SY, Chu CC. Down-regulation of *OsGRF1* gene in rice *rhd1* mutant results in reduced heading date. J Integr Plant Biol. 2005;47(6):745–52.

58. Choi D, Kim JH, Kende H. Whole genome analysis of the OsGRF gene family encoding plant-specific putative transcription activators in rice (*Oryza sativa* L.). Plant Cell Physiol. 2004;45(7):897–904.

59. Filiz E, Koc I, Tombuloglu H. Genome-wide identification and analysis of growth regulating factor genes in *Brachypodium distachyon*: in silico approaches. Turk J Biol. 2014;38(2):296–306.

60. Treich I, Cairns BR, Delossantos T, Brewster E, Carlson M. Snf11, a new component of the yeast Snf-Swi complex that interacts with a conserved region of Snf2. Mol Cell Biol. 1995;15(8):4240–8.

61. Yu J, Jung S, Cheng CH, Ficklin SP, Lee T, Zheng P, Jones D, Percy RG, Main D. CottonGen: a genomics, genetics and breeding database for cotton research. Nucleic Acids Res. 2014;42(Database issue):1229–36.

62. Wong DC, Schlechter R, Vannozzi A, Holl J, Hmmam I, Bogs J, Tornielli GB, Castellarin SD, Matus JT. A systems-oriented analysis of the grapevine R2R3-MYB transcription factor family uncovers new insights into the regulation of stilbene accumulation. DNA Res. 2016.

63. Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. 2013;41(12):e121.

64. Poole RL. The TAIR database. Methods Mol Biol. 2007;406:179–212.

65. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493–6.

66. Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 2016;44(D1): D279–85.

67. Voorrips RE. MapChart: software for the graphical presentation of linkage maps and QTLs. J Heredity. 2002;93(1):77–8.

68. 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.

69. Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. Methods Mol Biol. 1999;112:531–52.

70. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30(4):772.

71. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28(10):2731–9.

72. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60.

73. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc. 2016;11(9):1650–67.

74. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014;32(9):896–902.

75. Wang M, Wang Q, Zhang B. Evaluation and selection of reliable reference genes for gene expression under abiotic stress in cotton (*Gossypium hirsutum* L.). Gene. 2013;530(1):44–50.

76. Zhang Y, Parmigiani G, Johnson WE. Combat-seq: batch effect adjustment for RNA-seq count data. *bioRxiv*. 2020; 2020.2001.2013.904730.

77. Saeed AI, Sharov V, White J, Li J, Liang W, Bhagabati N, Braisted J, Klapa M, Currier T, Thiagarajan M, et al. TM4: a free, open-source system for microarray data management and analysis. BioTechniques. 2003;34(2):374–8.

78. Shan CM, Shangguan XX, Zhao B, Zhang XF, Chao LM, Yang CQ, Wang LJ, Zhu HY, Zeng YD, Guo WZ, et al. Control of cotton fibre elongation by a homeodomain transcription factor GhHOX3. Nat Commun. 2014;5:5519.

79. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(−Delta Delta C) method. Methods. 2001;25(4):402–8.

Publisher's Note