BMC Genomics

Open Access

CrossMark

Genome-wide analysis of GRAS transcription factor gene family in *Gossypium hirsutum* L.

Bin Zhang[1,2], J. Liu[2], Zhao E. Yang[2], Er Y. Chen[2], Chao J. Zhang[2], Xue Y. Zhang[2*] and Fu G. Li[1,2*]

Abstract

Background: Cotton is a major fiber and oil crop worldwide. Cotton production, however, is often threatened by abiotic environmental stresses. GRAS family proteins are among the most abundant transcription factors in plants and play important roles in regulating root and shoot development, which can improve plant resistance to abiotic stresses. However, few studies on the GRAS family have been conducted in cotton. Recently, the *G. hirsutum* genome sequences have been released, which provide us an opportunity to analyze the GRAS family in *G. hirsutum*.

Results: In total, 150 GRAS proteins from *G. hirsutum* were identified. Phylogenetic analysis showed that these GRAS protins could be classified into 14 subfamilies including SCR, DLT, OS19, LAS, SCL4/7, OS4, OS43, DELLA, PAT1, SHR, HAM, SCL3, LISCL and G_GRAS. The gene structure and motif distribution analysis of the GRAS members in *G. hirsutum* revealed that many genes of the SHR subfamily have more than one intron, which maybe a kind of form in the evolution of plant by obtaining or losing introns. Chromosomal location and duplication analysis revealed that segment and tandem duplication maybe the reasons of the expension of the GRAS family in cotton. Gene expression analysis confirmed the expression level of GRAS members were up-regulated under different abiotic stresses, suggesting that their possible roles in response to stresses. What's more, higher expression level in root, stem, leaf and pistil also indicated these genes may have effect on the development and breeding of cotton.

Conclusions: This study firstly shows the comprehensive analysis of GRAS members in *G. hirsutum*. Our results provide important information about GRAS family and a framework for stress-resistant breeding in *G. hirsutum*.

Keywords: *Gossypium hirsutum* L., GRAS family, Abiotic stress, Phylogeny, Duplication

Background

Transcription factors are proteins that play important roles in plant growth, development and resistance to abiotic environmental stresses, such as cold, hot, drought and salt. In general, transcription factors function by combining with other proteins or DNA sequences to promote or inhibit gene transcription. Some of the known families of transcription factors including WRKY, MYB, MADS, ARF, AP2/EREBP, HB, SBP, bZIP and GRAS families, have been well studied [1–3]. The GRAS transcription factor family is a plant-specific gene family named after the first three functionally characterized genes, GAI [4], RGA [5], and SCR [6]. And GRAS family likely emerged first in bacteria [7]. Generally, GRAS proteins range from 400 to 700 amino acid residues [8]. According to previous reports, GRAS proteins can be divided into at least 13 subfamilies including SCR, DLT, OS19, LAS, SCL4/7, OS4, OS43, DELLA, PAT1, SHR, HAM, SCL3 and LISCL [8]. GRAS proteins contain a conserved sequence in their carboxyl terminus, which typically contains several ordered motifs including LHRI, VHIID, LHRII, PFYRE and SAW [9]. Up to now, GRAS transcription factors have been found in more and more species, such as *Arabidopsis thaliana* [9], *Oryza sativa* [9], *Castor Bean* [8], tobacco [10], *Populus* [11], *Chinese cabbage* [12], *Grapevine* [13], *Zea mays L.* [14] and *Nelumbo nucifera* [15].

* Correspondence: zhangxueyan_caas@126.com; aylifug@163.com
[2]Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, China
[1]Research Base, Anyang Institute of Technology, State Key Laboratory of Cotton Biology, Anyang 455000, China

Zhang et al. BMC Genomics (2018) 19:348

Page 2 of 12

GRAS proteins play important roles in many processes such as signal transduction, axillary shoot meristem formation, root radial patterning, stress responses [16] and meristem maintenance [17]. The genes GAI, RGA, and RGL of DELLA subfamily were reported as repressors of gibberellin signaling [18] . Besides, previous reports confirmed that SCR and SHR could form the SCR/SHR complex, which was related with the root radial patterning [19] . In addition, SCR also controls Arabidopsis root meristem size from the root endodermis tissue by regulating the DELLA protein RGA [19]. What's more, SCL3 was proved as a factor mediating the elongation of root [20]. When the homolog of MOC1 in Arabidopsis, AtLAS, was knockout, forming lateral shoots during vegetative development became impossible [21]. In addition, the triple scl6 mutants performed similar pleiotropic phenotypes of meristematic cells when overexpressing miR171 [22]. Both the shr aba2 and scr aba2 double mutants of A. thaliana suppressed the effects of stresses on root meristematic activity [23]. Thus, GRAS transcription factors play important regulatory roles in the development of roots and shoots, which are always related with resistance to abiotic stresses. And several genes reported played a positive role in plant stress responses. The Arabidopsis GRAS protein SCL14 was shown to be essential for the activation of stress-inducible promoters [24]. What's more, NtGRAS1 in tobacco could raise the ROS level under various stresses [16]. Overexpressing PeSCL7 of P. euramaricana in Arabidopsis enhanced its tolerance to drought and salt treatments [25]. And OsGRAS23 was reported as a positive factor for stress resistance [26]. Besides, Overexpression of VaPAT1, a GRAS member from Vitis amurensis, confers abiotic stress tolerance in Arabidopsis [27].

Cotton is one of the most important fiber and oil crops worldwide but cotton production is always threatened by abiotic stresses, such as cold, heat, drought and salinity [28]. On one hand, breeding for stress-resistant cotton is a high priority for plant biotechnology programs. As plant-specific transcription factors, GRAS members play important roles in phytohormone pathways and the crosstalk among them [29]. DELLA-like genes have possible effect on cell elongation, influencing plant hight or root elongation, which may regulate the resistance to stresses [30]. Additionally, the GRAS family was enriched in the early response to salt stress in an analysis of serial cDNA libraries [31]. Therefore, studying the GRAS protein in cotton could help us breed stress-resistant cotton. Here, we identified 150 GRAS members in G. hirsutum. The phylogenetic relationships, locations, structures and expression patterns of these genes were analyzed. About 37 persents of members cotaining introns were members of SHR subfamily,

which maybe a kind of possible evolution pattern by obtaining or losing introns. And tandem should be the reason of the expansion of SHR family members. Ka/Ks of segmental gene pairs were lower than 1, suggesting that purified selection acted on these duplicated genes. In addition, the expression profiles of the GRAS members showed that they were differentially expressed under various stress treatments and among tissue. And we selected 15 genes up-regulated under stresses to confirm via qRT-PCR experiment, getting a similar result. Several genes could be good candidates for stress-resistant cotton breeding.
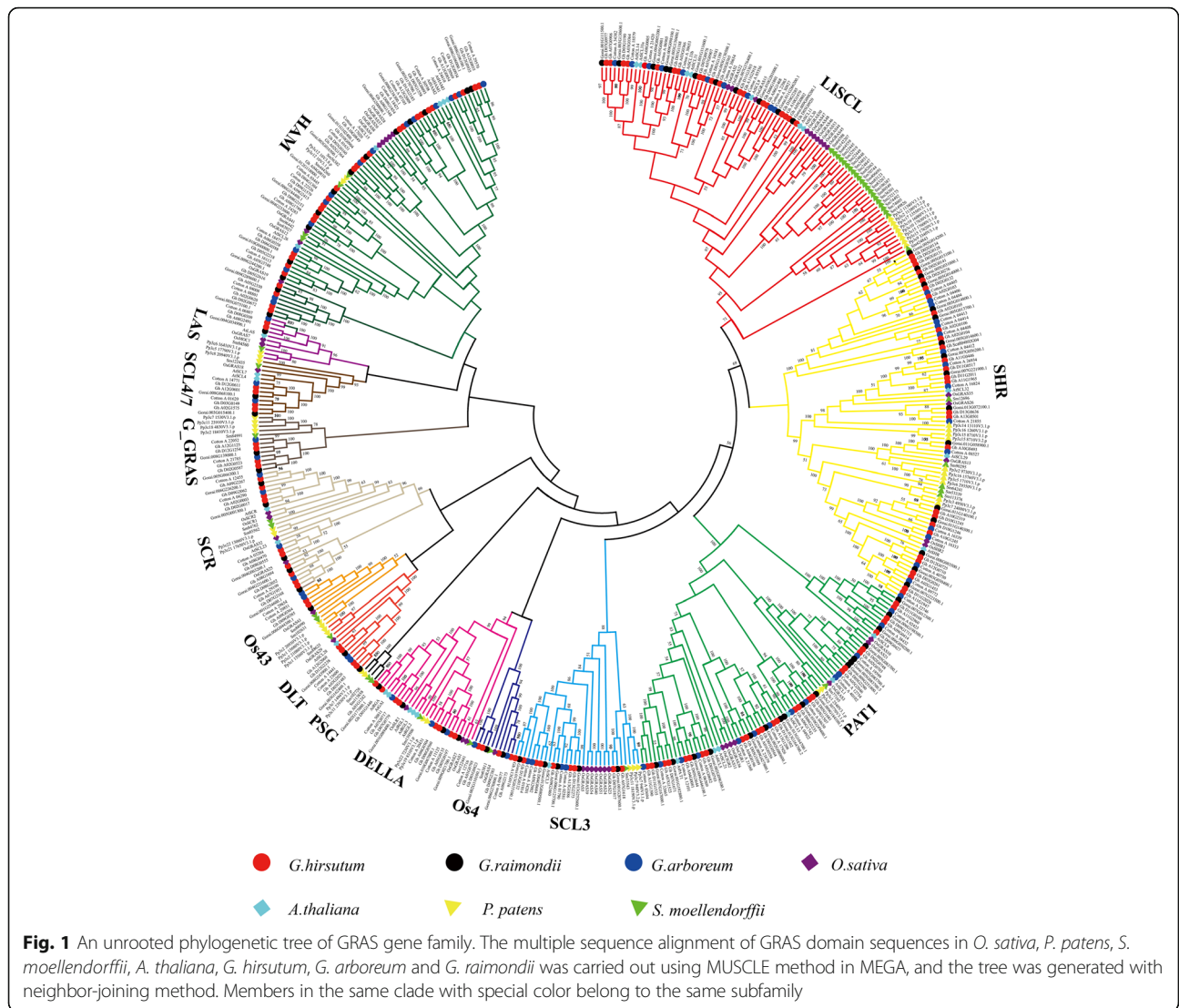
Results

Identification of GRAS family members

We identified a total of 150 GRAS proteins in G. hirsutum by a combination of methods. Additionally, 77 GRAS members in G. arboreum, 82 GRAS members in G. raimondii, 33 GRAS members in A. thaliana [9], 42 GRAS members in P. patens, 46 GRAS members in S. moellendorffii and 50 GRAS members in O. sativa [8] were identified with the same methods (Additional file 1: Table S1). The number of GRAS members in G. hirsutum (AD group) was about twice that in G. arboreum (A group) and G. raimondii (D group), which was consistent with their tetraploid and diploid genomes. Basic information, including gene length, subfamily, isoelectric point (pIs), and molecular weight (MWs), for the GRAS proteins in G. hirsutum is listed (Additional file 1: Table S1). Information on the GRAS proteins in the other species is provided (Additional file 1: Table S1). The identified GRAS encoded proteins ranging from 208 (Gh_D04G1304) to 1821 (Gh_D02G0276) amino acids, with pIs varying from 4.68 (Gh_A09G2415) to 8.99 (Gh_A05G0802) and MWs varying from 23,116.5 D (Gh_D04G1304) to 208,508.85 D (Gh_D02G0276). And, 141 members can be located on certain chromosomes, 67 members in the A group chromosomes and 74 members in D group chromosomes.

Phylogenetic analysis of GRAS members

We built a NJ phylogenetic tree (Fig. 1) to study the deeper relationships among the GRAS family members in cotton. On the basis of previous GRAS family studies [8], we divided all GRAS members identified into 15 subfamilies: SCR, DLT, OS19, LAS, SCL4/7, OS4, OS43, DELLA, PAT1, SHR, HAM, SCL3, LISCL, G_GRAS and PSG. G_GRAS is cotton-special subfamily, and PSG only contains genes of P. patens and S. moellendorffii. Most of the GRAS members enrich in LISCL, SHR, PAT1 and HAM. The distribution of genes of G. hirsutum among different subfamilies was as following: SCR (10), DLT (4), OS19 (4), LAS (2), SCL4/7 (4), OS4 (4), OS43 (2), DELLA (8), PAT1 (27),

Zhang *et al. BMC Genomics* (2018) 19:348

Page 3 of 12

Fig. 1 An unrooted phylogenetic tree of GRAS gene family. The multiple sequence alignment of GRAS domain sequences in *O. sativa*, *P. patens*, *S. moellendorffii*, *A. thaliana*, *G. hirsutum*, *G. arboreum* and *G. raimondii* was carried out using MUSCLE method in MEGA, and the tree was generated with neighbor-joining method. Members in the same clade with special color belong to the same subfamily

SHR (28), HAM (24), SCL3 (9), LISCL (20) and G_ GRAS (4). To check the results obtained by the NJ method, we constructed phylogenetic trees using the UPGMA evolution methods (Additional file 2: Figure S1). The final trees constructed with the two methods were almost same except two clades marked with imaginary lines (Additional file 2: Figure S1). Here, we identified GRAS members in moss (*P. patens*), fern (*S. moellendorffii*), dicotyledon (*G. hirsutum, G. raimondii, G. arboreum and A. thaliana*) and monocotyledon (*O. sativa*), finding there were GRAS members of moss and fern in every subfamily. This suggests that GRAS subfamilies occurred before these species. We could not, however, identified any GRAS members in almost all algaes, indicating that GRAS family may produce by bacterial infection [7] in the period between algae and moss. Finally, we build a model to describe the possible evolution process (Additional file 3: Figure S6).

Chromosomal locations and duplication of GRAS members

Analysis of the *G. hirsutum* genome sequence with the GFF3 profile revealed the chromosomal position of each GRAS member. As a result, 141 GRAS transcription factor genes were mapped onto the *G. hirsutum* chromosomes named according to their chromosomal order, A01–A13 and D01–D13. However, 9 GRAS members were not obviously mapped to any chromosome and were located on unattributed scaffolds (Additional file 1: Table S1). It indicates that genes' distribution on A chromosomes was almost same with the situation D chromosomes (Additional file 4: Figure S2).

Gene duplication analysis revealed that five kinds of duplication events (singleton, dispersed, proximal, tandem and segmental) (Additional file 5: Table S4) were reasons for GRAS family expansion in the cotton

Zhang *et al. BMC Genomics* (2018) 19:348

Page 4 of 12

genome. We identified 11 genes belonging to tandem duplications, 41 genes belonging to singleton duplications, 21 genes belonging to dispersed duplications, 4 genes belonging to proximal duplications and 64 genes belonging to segmental duplications, which showed that segmental duplication events were the main reason for GRAS members' evolution in *G. hirsutum*. In addition, homologous gene among GRAS members and segmental genes' pairs were showed by marking with yellow and blue lines respectively (Fig. 2). And we also calculated the Ka and Ks of segmental genes' pairs, finding Ka/Ks were all lower than 1, which indicated that purified selection happened on these genes. Finally, we identified 6 pairs of tandem duplicated gene pairs (Additional file 5: Table S4).

Exon-intron structure and domain architecture of GRAS members

Using the online software Gene Structure Display Server (GSDS 2.0, http://gsds.cbi.pku.edu.cn/index.php), we generated a graph to show the exon-intron

structures and GRAS domain positions of every GRAS member (Fig. 3c). And it clearly shows there was a conserved motif in the same subfamily. And the *p-values* of motifs showed on every protein were lower than $1e^{-5}$ (Fig. 3b). Besides, almost 30% members containing introns belong to SHR. And the related motifs were also different with others, whatever the length and motif orders. All motifs' logo were also showed (Additional file 6: Figure S4). We also could find that the members cotaining many introns lack the motif 17, which is very conservative in other proteins. Several special motifs (motif 11, motif 15 and motif 18) arise in these proteins.

According to previous studies, some GRAS members are regulated by microRNA 171 [32]. Thus, we queried the previously reported binding site sequence targeted by gh-microRNA 171 in cotton [33] for every member of the GRAS family. The results indicated that only 10 genes from the HAM subfamily (Fig. 4) were unambiguously complementary to the gh-microRNA 171 sequence.
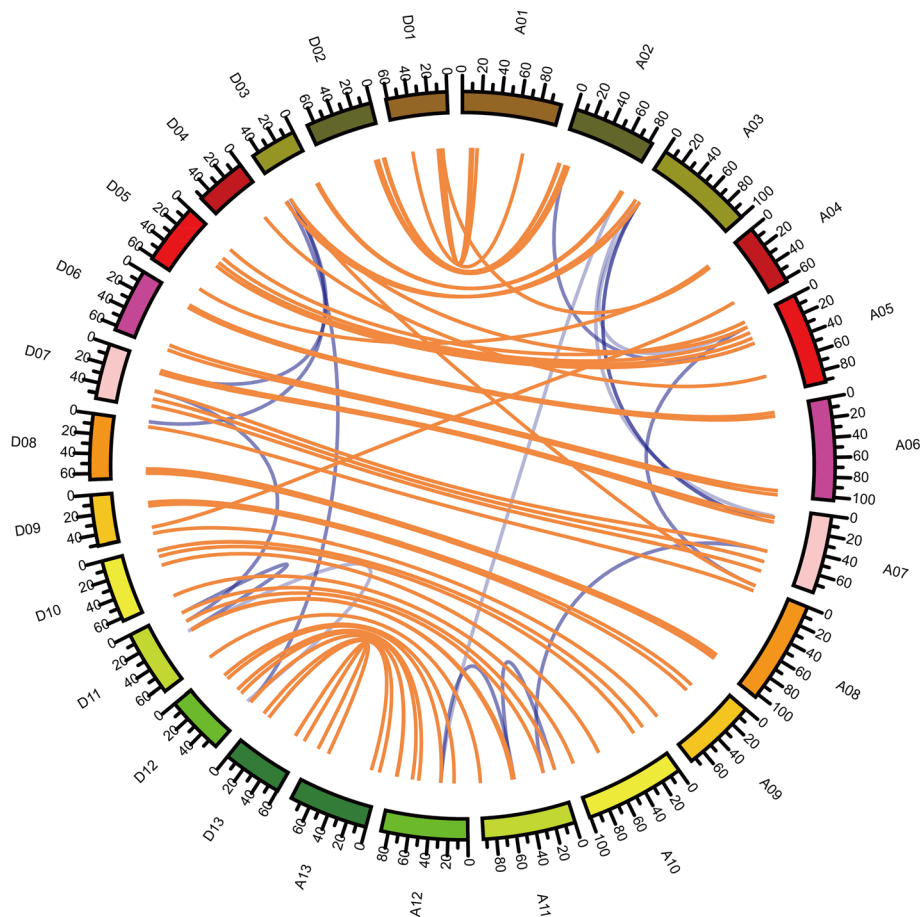

Fig. 2 Duplication events among GRAS members in *G. hirsutum* L.. Yellow lines indicate homologous genes between A and D chromosomes, and blue lines indicate segmental duplications between A or D chromosomes
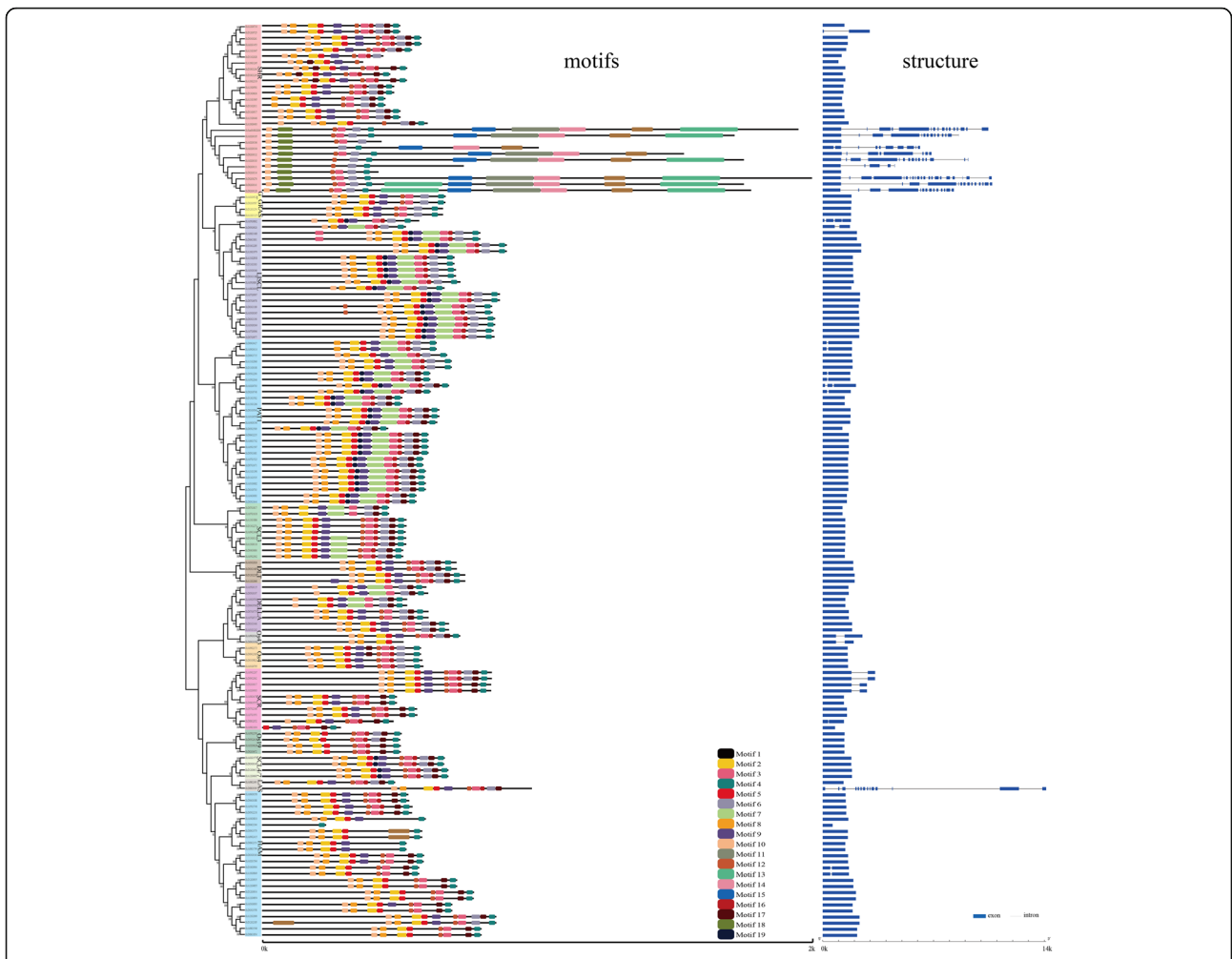
Zhang *et al. BMC Genomics* (2018) 19:348

Page 5 of 12

Fig. 3 Phylogenetic relationship, motifs and gene structures of GRAS members in *G. hirsutum* L.

Fig. 4 Potential GRAS members targeted by gh-miRNA 171 in *G. hirsutum* L.

Zhang et al. BMC Genomics (2018) 19:348

Page 6 of 12

Expression profiles of GRAS members under abiotic stresses, among different tissues and development stages

To investigate the expression profiles of the *G. hirsutum* GRAS members under abiotic stresses (heat, cold, salt and drought), we used publicly available transcriptome data [34]. Among several subfamilies (LISCL, HAM, PAT1 DELLA, SCL4/7 and SCR), genes had higher expression levels under stresses comparing with CK (Fig. 5b). In order to show the variation trend directly, several broken line graphs were built (Additional file 7: Figure S7). Additionally, we found that many genes were indeed induced by stress treatments (Additional file 7: Figure S7). Besides, in order to study the possible reasons of stress resistance, we also analyzed the expression pattern in different tissues (Fig. 5c) and development stages (Additional file 8: Figure S5). Most of *G. hirsutum* GRAS members have higher expression level in root, stem, leaf and pistil (Fig. 5b). Besides, about 70% genes in LISCL and PAT1 have higher expression level in different development stages (Additional file 8: Figure S5), suggesting these members maybe house-keeping genes.

qRT-PCR on 15 GRAS members under abiotic stresses

To validate the publicly available data downloaded from NCBI, 15 GRAS members whose expression level were up-regulated under stress treatments. We designed primers for 15 selected genes (Additional file 9: Table S2) and conducted a qRT-PCR experiment. According to the results, the expression profiles (Additional file 10: Table S5) of most selected genes were basically consistent with the published data, indicating the FPKM values (Additional file 11: Table S3) was reliable (Fig. 6). As the situation shown, *Gh_D02G0141, Gh_A06G1468, Gh_A03G0366, Gh_D03G1189, Gh_A05G2046, Gh_D01G0564, Gh_A04G0196, Gh_A01G0682* and *Gh_A04G0081* were significantly up-regulated under salt and PEG stresses ($P < 0.01$). What's more, *Gh_D02G0141* and *Gh_A04G0081* were all significantly up-regulated under hot ($P < 0.01$), which is same as the phenomenon shown by Zhang [35] (Fig. 5b). As for cold, *Gh_D02G0141, Gh_A03G0366, Gh_D03G1189, Gh_D01G0564* and *Gh_A04G0196* were up-regulated ($P < 0.01$). What valuable is that *Gh_D02G0141* have obvious resistance against four abiotic stresses, suggesting that this gene should work as the candidate for breeding stress-resistance cotton.

Discussion

GRAS family members have diverse functions and play important roles in plant development and physiological processes, such as GA signal transduction, root development, axillary shoot development, and transcriptional regulation in response to abiotic and biotic stresses [35]. To date, many studies have been undertaken to elucidate the evolutionary history of the GRAS family in a number of plant species, including *A. thaliana* [9], *O. sativa* [9], *Castor Bean* [8], tobacco [10], *Populus* [11], Chinese cabbage [12], Grapevine [13] and *Jatropha curcas* [36]. Nevertheless, few studies have been conducted in cotton. In this study, we performed a comprehensive analysis of the GRAS members in cotton, including their phylogenetic relationships, gene structures, conserved motifs, chromosomal locations, duplication events, and expression profiles under abiotic stresses.

Phylogeny, exon-intron structures and expansion of GRAS member in cotton

In this study, we divided the GRAS members identified in *O. sativa, A. thaliana, P. patens, S. moellendorffii*, and cotton (*G. hirsutum, G. arboreum* and *G. raimondii*) into 15 subfamilies. According to our results, most GRAS members (82%) have no intron. Interestingly, many SHR subfamily members blended with introns nabbed in *G. hirsutum*, suggesting these genes may be more recently evolved GRAS members in cotton. In order to explain this hypothesis, we searched the GRAS members range from algae to monocotyledon. We did not find any GRAS members in several kinds of algae (Additional file 3: Figure S6). However, there were many GRAS members in other plants. In consideration previous report on deriving from bacteria [7], we speculated that the origin of GRAS family may happen by the infection of bacteria in stage between algae and moss. What's more, we find there were many introns in *P. patens* and cotton, mainly enriching in SHR subfamily (Additional file 12: Figure S3). There existed, however, few intron among other species or subfamilies. According to previous reports, introns may play a role in the evolution of plants [37]. In the earlier stage of expansion, there always raised a lot of introns, losing slowly as time went by [38]. So, the more advanced of the specie, the fewer intron among genome [39]. Sometimes, however, introns also could become larger in the process of evolution, whose reason maybe the necessity for new function [40]. Besides, tandem duplication always resulted in the increasing of introns, leading to the generation of new genes [41]. In this study, cotton's introns increased which mainly belong to SHR, suggesting these genes may arise before the differentiation of monocotyledon and dicotyledon. What's more, 50% tandem duplicated gene pairs (Gh_A02G0103, Gh_A02G0104; Gh_A02G0105, Gh_A02G0106; and Gh_A02G0104, Gh_A02G0105) belong to SHR subfamily and are located in chromosome A02, indicating tandem duplication lead to the increasing of introns and expansion of genes. Furthermore, we can speculate the different motif among these genes that have many introns may possess special function for cotton's growth and development, which need deeper researches.
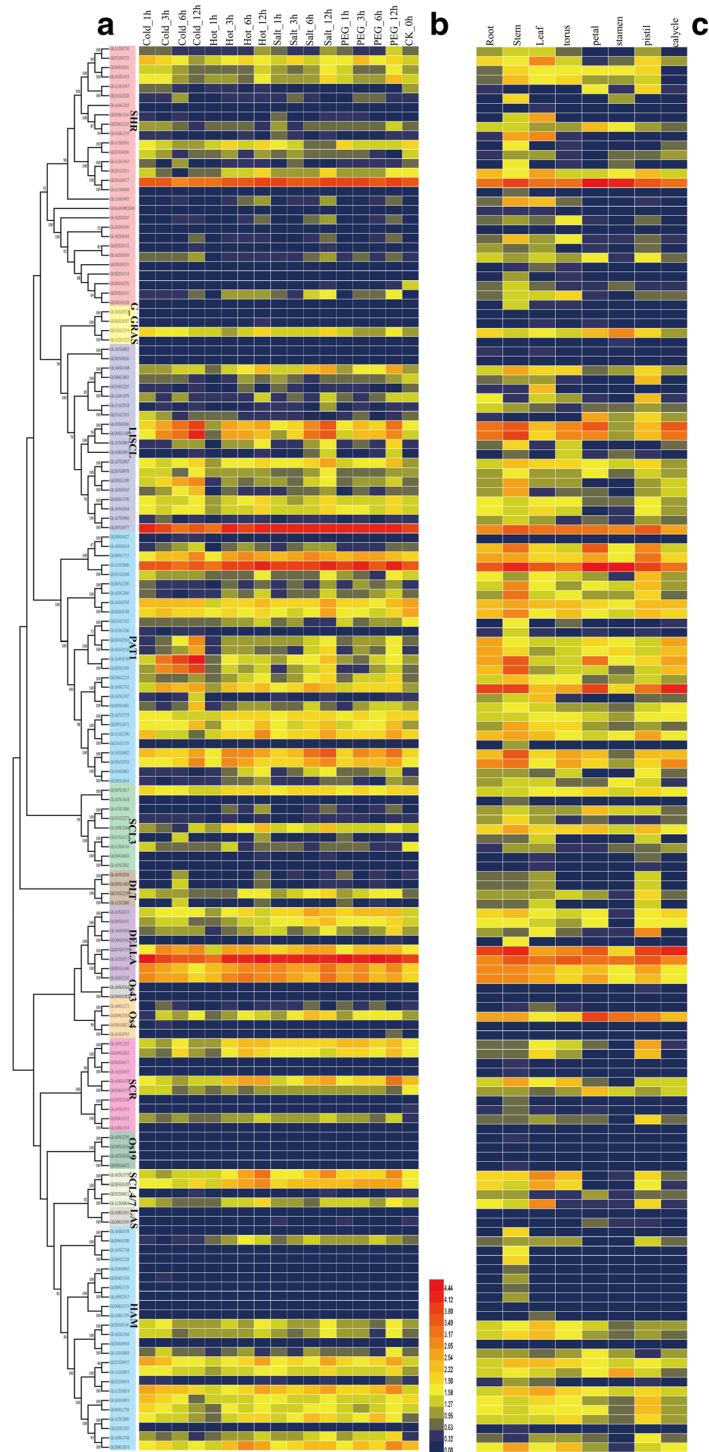
Zhang *et al. BMC Genomics* (2018) 19:348

Page 7 of 12

Fig. 5 Gene expression profiles of GRAS members in *G. hirsutum* L.. **a** A tree of the GRAS members in *G. hirsutum* L. **b** Gene expression profiles of GRAS members under different abiotic stresses. **c** Gene expression profiles of GRAS members in different tissues

The model for the evolution of GRAS family

Additionally, we proposed a model for the evolution of this gene family (Additional file 3: Figure S6). Subfamilies including SCR, DLT, Os19, LAS, SCL4/7, Os43, DELLA, PAT1, SHR, HAM, SCL3, LISCL, G_GRAS and PSG origin after the alga species. And then, Os4 was origin after the bryophyte. PSG, however, lost after the fern. What's more, some families including Os4, Os43 and G_
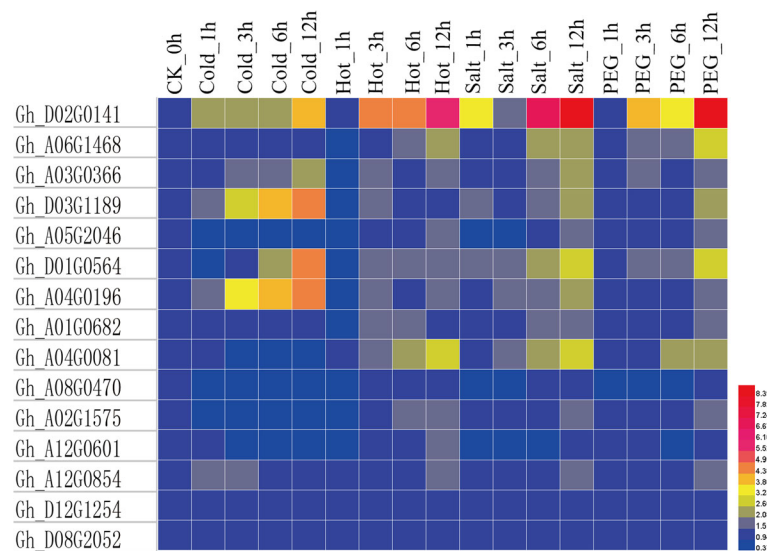
Zhang *et al. BMC Genomics* (2018) 19:348

Page 8 of 12

Fig. 6 The examination results of 15 selected GRAS members via qRT-PCR

GRAS loss in some dicots (such as *A. thaliana*). And subfamily G_GRAS lost after the dicots. What mentioned above shows that most subfamilies origin in the beginning. It also indicated that G_GRAS genes may have a special effect on the growth and development of cotton.

Analysis of the possible functions of the GRAS members

Importantly, the DELLA motif is required for the GA-induced degradation of DELLA proteins [42, 43]. Here eight DELLA genes were identified (Gh_A01G1242, Gh_A05G0135, Gh_A06G0504, Gh_A07G0717, Gh_D01 G1446, Gh_D05G0197, Gh_D06G0388 and Gh_D06 G2235), which possess DELLA domain and play important roles in regulation of the elongation of roots. In addition, previous reports have shown that microRNA 171 can regulate the expression of some GRAS members [44] Overexpressing miR171 had effect on plant height, flowering time, leaf architecture, phase transitions and floral meristem determinacy [45]. In this study, to our surprise, 10 GRAS members belonging to the HAM subfamily contained potential target sites for *gh-miRNA 171* binding, indicating a possible regulatory mechanism that needs to be studied in detail. Additionally, about 50% HAM subfamily members had higher expression levels under various abiotic stresses and among tissues (Fig. 5), suggesting that *gh-miRNA 171* may regulate pleiotropic phenotypes. So, this result would be helpful for us to study its signal path on abiotic stresses or generation.

Expression profiles of GRAS members in cotton

Many reports have provided evidence that GRAS members are important for plants to regulate their biochemical activities under various abiotic stresses [46].

According to data from a previous study [34], about 91% of the GRAS members identified in *G. hirsutum* are affected by a variety of stress conditions. Previous studies has shown that *NtGRAS1* [47] and *AtSCL14* were stress-induced members which belonged LISCL subfamily. Here, we found that *Gh_D02G0141, Gh_A03G0366, Gh_D03G1189, Gh_D01G0564* and *Gh_A04G0196* were obviously up-regulated when being under the cold, salt and PEG treatments. The expression level of *Gh_A04G0081*, did not change under cold, but be up-regulated under other three kinds of stressess. In addition, *Gh_D01G0564* and *Gh_A04G0196* belonging to PAT1 subfamily were up-regulated (*P < 0.01*) whatever the stresses, inducating this gene may play important roles in anti-stress for cotton, which show no difference with the report on *VaPAT1* [27]. Over-expression of *Gh_GRAS42* may improve cotton's tolerance to abiotic stresses. What's more, Ma [25] has proved that *PeSCL7*, the homologous gene of *AtSCL7* of SCL4/7 subfamily member, could be induced by salt and drought stresses, which was helpful for *Arabidopsis thaliana* to improve salt and drougt tolerance. We also found that SCL4/7 subfamily member *Gh_A02G1575* was up-regulated (*P < 0.01*) when treated under PEG and salt (Fig. 5b), which show it is important for cotton to resist abiotic stresses. What's more, its expression level was higher among root, stem, leaf and pistil, suggesting it may also influenced development and generation (Fig. 5c). Interestingly, *Gh_D02G0141*, whose expression level was clearly high under all kinds of stresses (*P < 0.01*), indicating that this gene may be a good gene for breeding cotton resisting abiotic stresses. In addition, 50–70% genes of LISCL, PAT1 and DELLA have continual higher expression level in different development

Zhang et al. BMC Genomics (2018) 19:348

Page 9 of 12

stages, showing these genes maybe house keeping genes (Additional file 8: Figure S5B). Together, our results show that G. hirsutum GRAS members are necessary for abiotic stress resistance and could be used to improve the stress tolerance of cotton in future studies.

Conclusions

In this study, 150 GRAS members in G. hirsutum, 77 GRAS members in G. arboreum, 82 GRAS members in G. raimondii, 33 GRAS members in A. thaliana [9] 42 GRAS members in P. patens, 46 GRAS members in S. moellendorffii, and 50 GRAS members in O. sativa [8] were identified. The GRAS family was divided into 5 subgroups based on the phylogenetic tree and the distribution of conserved motifs. Intron/exon structure and motif analysis showed GRAS members may produce some genes cotaining introns, whose expansion may because tandem duplication. The chromosomal location of every GRAS member was identified, and duplication analysis showed that segmental duplication events played the main role in GRAS family expansion. And purified selection worked on these duplicated pairs. Additionally, we also found that gh-miRNA 171, which plays a regulatory role in GRAS member expression according to previous reports, had target genes in G. hirsutum, all of which belonged to the HAM subgroup. Finally, the expression patterns of the GRAS members under abiotic stresses and among tissues or development stages were explored using transcriptome data and qRT-PCR. In conclusion, our work provides functional insights into the roles of GRAS members in the resistance of cotton to abiotic stresses.

Methods

Identification of GRAS family members in cotton and other species

We downloaded the latest versions of the genome annotations for the following species from the Phytozome v11 database (https://phytozome.jgi.doe.gov/pz/portal.html): C. reinhardtii (V5.5), P. patens (V3.3), S. moellendorffii (V1.0), A. thaliana (TAIR annotation release 10) and O. sativa (MSU annotation release 7.0). Additionally, we downloaded the G. hirsutum [34] (https://www.cottongen.org/species/Gossypium_hirsutum/nbi-AD1_genome_v1.1) , Gossypium arboreum [48] (https://www.cottongen.org/species/Gossypium_arboreum/bgi-A2_genome_v2.0) and Gossypium raimondii [49] (https://www.cottongen.org/species/Gossypium_raimondii/jgi_genome_221) genome sequences from COTTONGEN (https://www.cottongen.org/). We download the hidden Markov model (HMM) profile (PF03514.13) [11]. The HMM profile was then used to perform HMM searches against annotated protein databases from different genomes with an E-value cutoff

of $1e^{-5}$ in the program HMMER 3.1b2 [50]. To confirm that the predicted proteins contained GRAS domains, we evaluated them using the SMART software [51] (http://smart.embl-heidelberg.de/). Proteins without GRAS domain or the length were shorter than half of the typical GRAS domain length (350aa) [46]. And at the same time, we also recorded the GRAS domain positions so that we could retrieve the domain sequences for next study. What's more, we arranged GRAS proteins's pIs, lengths, weights, chrs, related genes' positions and subfamilies.

Phylogenetic analysis

The GRAS domain amino acid sequences were determined using the complete GRAS protein sequences identified by HMM profile searches and the domain positions identified by the SMART software. The GRAS proteins' domain sequences in A. thaliana, O. sativa, C. reinhardtii, P. patens, S. moellendorffii and cotton were aligned using muscle method of MEGA6.0 with the default parameters. The aligned sequences were then subjected to phylogenetic analysis with the neighbor joining (NJ) method. The consensus tree was built using MEGA 6.0 [52] under the P-distance model and the pairwise deletion option with 1000 bootstrap replicates. Additionally, we used the UPGMA method to examine the results from the NJ method. Finally, we divided subfmily according previous reports and modified the phylogenetic trees by AI software. PoGOs were defined as described in Del Bem and Vincentz [53].

Analysis of the conserved motifs and gene structures

The online software Gene Structure Display Server (GSDS 2.0) (http://gsds.cbi.pku.edu.cn/index.php) [54] was used to show the gene structures (introns and exons) by submitting the related profile extracted from the genome's GFF3 profile (ftp://ftp.bioinfo.wsu.edu/species/gossypium_hirsutum/NAU-NBI_G.hirsutum). Then we performed motifs' analysis by local MEME software with following parameters: minsites 6, maxsites 149, minw 6, maxw 200 and nmotifs 19.

Chromosomal locations analysis of GRAS members

The chromosomal positions of all GRAS members were confirmed using the cotton genome annotation file (ftp://ftp.bioinfo.wsu.edu/species/Gossypium_hirsutum/NAU-NBI_G). The length of each chromosome was calculated using a Perl script. Then, we made a diagram of the GRAS members that showed duplications among 26 chromosomes and identified five types of gene duplication events with MCScanX using the GRAS protein sequences and the position data in chromosomes [55]. Additionally, we made a diagram of the chromosomal locations with the MapInspect software [56].

Zhang *et al. BMC Genomics* (2018) 19:348

Page 10 of 12

Calculation of Ka/Ks values

The protein sequences from segmentally duplicated pairs and orthologous pairs were aligned using Clustal X 2.0. Then, the aligned sequences were converted to the original cDNA sequences by the PAL2NAL program [57] (http://www.bork.embl.de/pal2nal/). Finally, The CODE ML program of the PAML package [58] was used to estimate the synonymous (Ks) and nonsynonymous (Ka) substitution rates.

Expression analysis of GRAS members

The expression levels of GRAS members were measured using the RNA-sequencing data of *G. hirsutum* TM-1 among different tissues, development stages and stresses (cold, heat, salt and drought), which were downloaded from the NCBI Sequence Read Archive (SRA: PRJNA248163) (http://www.ncbi.nlm.nih.gov/bioproject/PRJNA248163/) [59] The FPKM values (Additional file 11: Table S3) were performed treament of $\ln(X + 1)$, whose results were than used for building heatmap. Finally, a heatmap of GRAS expression was produced using the HemI 1.0 software [60].

RNA isolation and the qRT-PCR analysis

In order to validate the expression data we downloaded and prepare for further study, we selected 15 genes whose expression levels were obviously higher under stresses. The qRT-PCR primers were designed (Additional file 9: Table S2) and seedlings of CCRI24 was used. After 4 weeks of cultivation, plants at the three-leaf stage were subjected to four kinds of abiotic stress. The treatments were as follows: cold (4 °C for 1,3,6,12 h), heat (38 °C for 1,3,6,12 h), salt (400 mM NaCl for 1,3,6,12 h), and drought (20% PEG for 1,3,6,12 h) [34]. Samples collected at 0 h were regarded as controls. Ten plants were included in each treatment group and their leaves were collected at the three-leaf stage.

RNA was extracted from the detached leaves using an RNA Aprep Pure Plant Kit (TIANGEN, Beijing, China). The quality and concentration of each RNA sample was determined using gel electrophoresis and a NanoDrop 2000 spectrophotometer (only RNA that met the criteria A260/280 ratio = 1.8–2.1 and A260/230 ratio ≥ 2.0 was used for further analyses and stored at − 80 ℃). The high-quality RNA samples were treated with DNase I (TaKaRa, TOYOBO, Japan) to eliminate contaminating genomic DNA. cDNA was synthesized from 1 μg of RNA in a 20 μL reaction volume using the PrimeScript™ RT Reagent Kit (Takara, Dalian, China) according to the manufacturer's manual. qRT-PCR experiments were conducted to measure the expression levels of the GRAS members under stress treatments. qRT-PCR analysis was performed using the Applied Biosystems 7500 Real-Time PCR system and the SYBR premix Ex Taq™ kit (TaKaRa, Japan). Target gene amplification was checked by

the SYBR Green fluorescence signal. The cotton constitutive His gene [61] (GenBank accession no. AF024716) was used as a reference gene and 15 pairs of specific GRAS primers were used for qRT-PCR. Three biological replicates were included for each treatment, each consisting of three technical replicates.. The following thermal circulation conditions were used: 95 °C for 30 s, followed by 40 cycles of 95 °C for 5 s and 60 °C for 1 min. Expression levels were calculated as the mean signal intensity across the three replicates. Following the PCR, a melting curve analysis was performed. Ct or threshold cycle values were used for relative quantification of the input target number. The relative fold difference (N) is the number of treated target gene transcript copies relative to the untreated gene transcript copies, and was calculated according to Livak and Schmittgen (2001) as follows: $N = 2^{-\Delta\Delta Ct} = 2^{-(\Delta Ct \ \text{treated} - \Delta Ct \ \text{control})}$, where $\Delta\Delta Ct = \Delta Ct$ of the treated sample minus ΔCt of the untreated control sample. Student's t-test were conducted in order to study whether the variation of expression level under CK and treatments was meaningful.

Additional files

Additional file 1: Table S1. GRAS members identified in *O. sativa*, *P. patens*, *S. moellendorffii*, *A. thaliana*, *G. hirsutum*, *G. arboreum* and *G. raimondii*. (XLS 111 kb)

Additional file 2: Figure S1. Phylogenetic tree built with the UPGMA method. (TIF 8356 kb)

Additional file 3: Figure S6. Phylogenetic trees and evolutionary profile of GRAS genes in green plants. (A) Phylogenetic tree showing the evolutionary relationship between plants (marked with different colors) GRASSs. Tree topology is a consensus from NJ, UPGMA. Bootstrap values from the original trees higher than 50% are shown (NJ/UPGMA). (B) Evolutionary profile of GRAS subfamilyes' genes in green plants. (TIF 1487 kb)

Additional file 4: Figure S2. Chromosomal locations of GRAS members in *G. hirsutum* L.. (TIF 2000 kb)

Additional file 5: Table S4. The duplication types of GRAS members in *Gossypium hirsutum* L. and segmental gene pairs with their Ka/Ks values. Besides, six tandem duplicated gene pairs were also performed. (XLSX 14 kb)

Additional file 6: Figure S4. Motifs description of GRAS members in *G. hirsutum* L.. development stages in *G. hirsutum* L.. (TIF 6550 kb)

Additional file 7: Figure S7. The expression level of different GRAS genes under various stress. (TIF 1752 kb)

Additional file 8: Figure S5. Gene expression profile of GRAS members among different. (TIF 1197 kb)

Additional file 9: Table S2. Premiers of GRAS members selected for qRT-PCR. (XLS 22 kb)

Additional file 10: Table S5. The expression data of qRT-PCR. (XLS 36 kb)

Additional file 11: Table S3. The public available expression profile of GRAS members. (XLS 123 kb)

Additional file 12: Figure S3. The exon-intron structure of every GRAS member in six species. (A) *P. patens* (B) *O. sativa* (C) *S. moellendorffii* (D) *G. arboreum* (E) *A. thaliana* (F) *G. raimondii*. (TIF 1283 kb)

Abbreviations

D: Dalton; FPKM: Fragments Per Kilobase of exon model per Million mapped reads; G_GRAS: Gossypium GRAS members; GA: Gibberellic acid; GRAS: GAI,

Zhang *et al. BMC Genomics* (2018) 19:348

Page 11 of 12

Availability of data and materials
The data sets supporting the results of this article are included within the article and its additional files.

Authors' contributions
XYZ and FGL conceived and designed the experiments. BZ, ZEY and JL performed the experiments, prepared the materials and conducted the manuscript. EYC and CJZ participated in data analyses. All authors read and approved the final manuscript.

Ethics approval
The *Gossypium hirsutum* acc. CCRI24 seeds used in this study were supplied by Dr. Xiongming Du (National Mid-term Genebank for Cotton, Institute of Cotton Research of CAAS). Plants were grown in the growth room for sample collection. The research conducted in this study required neither approval from an ethics committee, nor involved any human or animal subjects.

Consent for publication
All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing financial interests.

Publisher's Note

References
1. Alves M, Dadalto S, Gonçalves A, De Souza G, Barros V, Fietto L. Plant bZIP transcription factors responsive to pathogens: a review. Int J Mol Sci. 2013; 14(4):7815–28.
2. Kaur S, Iquebal MA, Jaiswal S, Tandon G, Sundaram RM, Gautam RK, Suresh KP, Rai A, Kumar D. A meta-analysis of potential candidate genes associated with salinity stress tolerance in rice. Agri Gene. 2016;1:126–34.
3. Theißen G, Melzer R, Rümpler F. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. Development. 2016;143(18):3259–71.
4. Peng J, Carol P, Richards DE, King KE, Cowling RJ, Murphy GP, Harberd NP. The Arabidopsis GAI gene defines a signaling pathway that negatively regulates gibberellin responses. Genes Dev. 1997;11(23):3194.
5. Silverstone AL, Ciampaglio CN, Sun T. The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. Plant Cell. 1998;10(2):155–69.
6. Di LL, Wysockadiller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. The SCARECROW gene regulates an asymmetric cell division that is essential for generating the radial organization of the Arabidopsis root. Cell. 1996;86(3):423–33.
7. Zhang D, Iyer LM, Aravind L. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. Bioinformatics. 2012;28(19): 2407–11.
8. Xu W, Chen Z, Ahmed N, Han B, Cui Q, Liu A. Genome-wide identification, evolutionary analysis, and stress responses of the GRAS gene family in Castor beans. Int J Mol Sci. 2016;17(7):1004.
9. Tian C, Wan P, Sun S, Li J, Chen M. Genome-wide analysis of the GRAS gene family in rice and Arabidopsis. Plant Mol Biol. 2004;54(4):519–32.
10. Chen YQ, Tai SS, Wang DW, Ding AM, Sun TT, Wang WF, Sun YH. Homology-based analysis of the GRAS gene family in tobacco. Genet Mol Res. 2015;14(4):15188–200.
11. Liu X, Widmer A. Genome-wide comparative analysis of the GRAS gene family in Populus, Arabidopsis and Rice. Plant Mol Biol Report. 2014;32(6):1129–45.
12. Song X-M, Liu T-K, Duan W-K, Ma Q-H, Ren J, Wang Z, Li Y, Hou X-L. Genome-wide analysis of the GRAS gene family in Chinese cabbage (Brassica rapa ssp. pekinensis). Genomics. 2014;103(1):135–46.
13. Grimplet J, Agudelo-Romero P, Teixeira RT, Martinez-Zapater JM, Fortes AM. Structural and functional analysis of the GRAS gene family in grapevine indicates a role of GRAS proteins in the control of development and stress responses. Front Plant Sci. 2016;7:353.
14. Guo Y, Wu H, Li X, Li Q, Zhao X, Duan X, An Y, Lv W, An H. Identification and expression of GRAS family genes in maize (Zea mays L.). PLoS One. 2017;12(9):e0185418.
15. Wang Y, Shi S, Zhou Y, Zhou Y, Yang J, Tang X. Genome-wide identification and characterization of GRAS transcription factors in sacred lotus (Nelumbo nucifera). PeerJ. 2016;4:e2388.
16. Mayrose M, Ekengren SK, Melech-Bonfil S, Martin GB, Sessa G. A novel link between tomato GRAS genes, plant disease resistance and mechanical stress response. Mol Plant Pathol. 2006;7(6):593–604.
17. Bolle C. The role of GRAS proteins in plant signal transduction and development. Planta. 2004;218(5):683–92.
18. Stuurman J, Jäggi F, Kuhlemeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. Genes Dev. 2002;16(17):2213–8.
19. Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Blilou I, Scheres B, Benfey PN. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. Science (New York, NY). 2007;316(5823):421–5.
20. Heo JO, Estelle M. Funneling of gibberellin signaling by the GRAS transcription regulator scarecrow-like 3 in the Arabidopsis root. Proc Natl Acad Sci U S A. 2011;108(5):2166.
21. Greb T, Clarenz O, Schafer E, Muller D, Herrero R, Schmitz G, Theres K. Molecular analysis of the LATERAL SUPPRESSOR gene in Arabidopsis reveals a conserved control mechanism for axillary meristem formation. Genes Dev. 2003;17(9):1175–87.
22. Wang L, Mai YX, Zhang YC, Luo Q, Yang HQ. MicroRNA171c-targeted SCL6-II, SCL6-III, and SCL6-IV genes regulate shoot branching in Arabidopsis. Mol Plant. 2010;3(5):794–806.
23. Cui H. Killing two birds with one stone: transcriptional regulators coordinate development and stress responses in plants. Plant Signal Behav. 2012;7(6): 701–3.
24. Fode B, Siemsen T, Thurow C, Weigel R, Gatz C. The Arabidopsis GRAS protein SCL14 interacts with class II TGA transcription factors and is essential for the activation of stress-inducible promoters. The Plant Cell Online. 2008;20(11):3122–35.
25. Ma H-S, Liang D, Shuai P, Xia X-L, Yin W-L. The salt- and drought-inducible poplar GRAS protein SCL7 confers salt and drought tolerance in Arabidopsis thaliana. J Exp Bot. 2010;61(14):4011–9.
26. Kai X, Chen S, Li T, Ma X, Liang X, Ding X, Liu H, Luo L. OsGRAS23, a rice GRAS transcription factor gene, is involved in drought stress response through regulating expression of stress-responsive genes. BMC Plant Biol. 2015;15(1):141.
27. Yuan Y, Fang L, Karungo SK, Zhang L, Gao Y, Li S, Xin H. Overexpression of VaPAT1, a GRAS transcription factor from Vitis amurensis, confers abiotic stress tolerance in Arabidopsis. Plant Cell Rep. 2016;35(3):655–66.
28. Huang J, Pang C, Fan S, Song M, Yu J, Wei H, Ma Q, Li L, Zhang C, Yu S. Genome-wide analysis of the family 1 glycosyltransferases in cotton. Mol Gen Genomics. 2015;290(5):1805–18.
29. Heo JO, Chang KS, Kim IA, Lee MH, Lee SA, Song SK, Lee MM, Lim J. Funneling of gibberellin signaling by the GRAS transcription regulator scarecrow-like 3 in the Arabidopsis root. Proc Natl Acad Sci U S A. 2011;108(5):2166.
30. Wang L, Mu C, Du M, Chen Y, Tian X, Zhang M, Li Z. The effect of mepiquat chloride on elongation of cotton (Gossypium hirsutum L.) internode is associated with low concentration of gibberellic acid. Plant Sci. 2014;225(8):15.

Zhang *et al. BMC Genomics* (2018) 19:348

Page 12 of 12

31. Shi G, Guo X, Guo J, Liu L, Hua J. 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.

32. Ma Z, Hu X, Cai W, Huang W, Zhou X, Luo Q, Yang H, Wang J, Huang J. Arabidopsis miR171-targeted scarecrow-like proteins bind to GT cis-elements and mediate gibberellin-regulated chlorophyll biosynthesis under light conditions. PLoS Genet. 2014;10:e1004519.

33. Wang Q, Liu N, Yang X, Tu L, Zhang X. Small RNA-mediated responses to low- and high-temperature stresses in cotton. Sci Rep. 2016;6(1):35558.

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

35. Cenci A, Rouard M. Evolutionary analyses of GRAS transcription factors in angiosperms. Front Plant Sci. 2017;8:273.

36. Wu ZY, Wu PZ, Chen YP, Li MR, Wu GJ, Jiang HW. Genome-wide analysis of the GRAS gene family in physic nut (Jatropha curcas L.). Genet Mol Res. 2015;14(4):19211–24.

37. Roy SW. The evolution of spliceosomal introns: patterns, puzzles and progress. Nat Rev Genet. 2006;7(3):211.

38. Roy SW, Penny D. A very high fraction of unique intron positions in the intron-rich diatom Thalassiosira pseudonana indicates widespread intron gain. Molecular Biology & Evolution. 2007;24(7):1447.

39. Roy SW, Gilbert W. Complex early genes. Proc Natl Acad Sci U S A. 2005; 102(6):1986–91.

40. Owamoto M, Maekawa M, Saito A, Higo H, Higo K. Evolutionary relationship of plant catalase genes inferred from exon-intron structures: isozyme divergence after the separation of monocots and dicots. Theoretical & Applied Genetics. 1998;97(1–2):9–19.

41. Hu X, Worton RG. Partial gene duplication as a cause of human disease. Hum Mutat. 1992;1(1):3–12.

42. Ariizumi T, Steber CM. Seed germination of GA-insensitive sleepy1 mutants does not require RGL2 protein disappearance in Arabidopsis. The Plant Cell Online. 2007;19(3):791–804.

43. Zhang Y, Liu Z, Liu J, Lin S, Wang J, Lin W, Xu W. GA-DELLA pathway is involved in regulation of nitrogen deficiency-induced anthocyanin accumulation. Plant Cell Rep. 2017;36(4):557–69.

44. Qu L-J, Ma Z, Hu X, Cai W, Huang W, Zhou X, Luo Q, Yang H, Wang J, Huang J. Arabidopsis miR171-targeted scarecrow-like proteins bind to GT cis-elements and mediate gibberellin-regulated chlorophyll biosynthesis under light conditions. PLoS Genet. 2014;10(8):e1004519.

45. Curaba J, Talbot M, Li Z, Helliwell C. Over-expression of microRNA171 affects phase transitions and floral meristem determinancy in barley. BMC Plant Biol. 2013;13(1):6.

46. Niu Y, Zhao T, Xu X, Li J. Genome-wide identification and characterization of GRAS transcription factors in tomato (Solanum lycopersicum). Peerj. 2017; 5(11):e3955.

47. Czikkel BE, Maxwell DP. NtGRAS1, a novel stress-induced member of the GRAS family in tobacco, localizes to the nucleus. J Plant Physiol. 2007;164(9): 1220–30.

48. 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.

49. Yeh RF, Lim LP, Burge CB. Computational inference of homologous gene structures in the human genome. Genome Res. 2001;11(5):803–16.

50. Wheeler TJ, Eddy SR. Nhmmer: DNA homology search with profile HMMs. Bioinformatics (Oxford, England). 2013;29(19):2487.

51. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43(Database issue):257–60.

52. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013;30(12):2725–9.

53. Del Bem LE, Vincentz MG. Evolution of xyloglucan-related genes in green plants. BMC Evol Biol. 2010;10(1):341.

54. Hu B, Jin J, Guo A, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics (Oxford, England). 2015;31(8):1296.

55. Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 2012;40(7):e49.

56. Jia J, Zhao P, Cheng L, et al. MADS-box family genes in sheepgrass and their involvement in abiotic stress responses. BMC Plant Biol. 2018;18(1):42.

57. Zhang C, Wang J, Long M, Fan C. gKaKs: the pipeline for genome-level Ka/Ks calculation. Bioinformatics. 2013;29(5):645–6.

58. Lu J, Zheng J, Xu Q, et al. Adaptive evolution of the vertebrate skeletal muscle sodium channel. Genetics & Molecular Biology. 2011;34(2):323.

59. Yang Z, Qian D, Qin W, Yang Z, Yuan C, Lu L, Ge X, Zhang C, Wu Z, Li F. Genome-wide analysis of WOX genes in upland cotton and their expression pattern under different stresses. BMC Plant Biol. 2017;17(1):113.

60. Deng W, Wang Y, Liu Z, Cheng H, Xue Y. HemI: a toolkit for illustrating Heatmaps. PLoS One. 2014;9(11):e111988.

61. Xu Y, Wang J, Wang S, Wang J, Chen X. Characterization of GaWRKY1, a cotton transcription factor that regulates the Sesquiterpene synthase gene (+)-δ-Cadinene synthase-a. Plant Physiol. 2004;135(1):507.