

1    **Declarations**

2    **Ethics approval and consent to participate:**

3    Not applicable.

4

5    **Consent for publication:**

6    All authors have agreed to publish.

7

8    **Availability of data and materials:**

9    The *Sorghum bicolor* whole genome sequence information is from the Ensembl Genomes website  
10    (<http://ensemblgenomes.org/>). The *Sorghum bicolor* materials (Hongyingzi) used in this study were  
11    supplied by Prof. Cheng Jianping of Guizhou University. The datasets supporting the conclusions  
12    of this article are included in the article and its Additional files.

13

14    **Competing interests:**

15    The authors declare that they have no competing interests.

16

17    **Funding:**

18    This research was supported by the National Science Foundation of China (31560578, Cheng JP,  
19    <http://www.nsf.gov.cn>), Sichuan International Science and Technology Cooperation and Exchange  
20    Research and Development Project (2018HH0116, Yan J, <http://kjt.sc.gov.cn>), and Guizhou Science  
21    and Technology Support Project (2020Y125).

22

23    **Authors' contributions:**

24    YF planned and designed the research and analyzed the data. YF and HY wrote the manuscript. DL,  
25    LF and LC studied gene expression by qRT-PCR. AH identified the *S. bicolor GRAS* gene family  
26    and analyzed gene structure. GX and AH studied chromosome distribution and gene duplication,  
27    performed the syntenic analysis of *S. bicolor GRAS* genes. YF and X-bC analyzed the evolutionary  
28    relationship of *S. bicolor GRAS* genes with those in several other species. JC supervised the research.  
29    JY and JR revised the manuscript. YF and DL contributed equally to the manuscript. All authors  
30    read and approved the final manuscript.

31

32 **Acknowledgements:**

33 We thank all of the colleagues in our laboratory for providing useful discussions and technical  
34 assistance. We are very grateful to the editor and reviewers for critically evaluating the manuscript  
35 and providing constructive comments for its improvement.

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

55

56

57

58

59

60

Genome-wide identification and expression analysis of the GRAS transcription factor family and its response to abiotic stress in sorghum [*Sorghum bicolor* (L.) Moench]

Yu Fan<sup>1†</sup>, Jun Yan<sup>2</sup>, Dili Lai<sup>1†</sup>, Hao Yang<sup>1</sup>, Guoxing Xue<sup>1</sup>, Ailing He<sup>1</sup>, Liang Feng<sup>3</sup>, Long Chen<sup>4</sup>, Xiao-bin Cheng<sup>5</sup>, Jingjun Ruan<sup>1\*</sup>, Jianping Cheng<sup>1\*</sup>

1 College of Agriculture, Guizhou University, Guiyang 550025, P.R. China

2 School of Pharmacy and Bioengineering, Chengdu University, Chengdu 610106, P.R. China

3 Chengdu Food and Drug Inspection Institute, Chengdu 610000, P.R. China

4 Department of Nursing, Sichuan Tianyi College, Mianzhu 618200, P.R. China

5 Department of Environmental and Life Sciences, Sichuan MinZu College, Kangding 626001, P.R. China

† Equal contribution.

**\* Corresponding author:** Jian-ping Cheng

Guizhou University, Huaxi District, Guiyang City, Guizhou Province, P.R. China

Tel.: 0851-88305238; Email: [chengjianping63@qq.com](mailto:chengjianping63@qq.com)

## Abstract

**Background:** GRAS is a very important family of transcription factors that are unique to plants, playing important roles in plant development and their response to abiotic stress. Since the sequencing of the sorghum genome, a large number of genetic studies based on this genomic information have been carried out. Nevertheless, no detailed identification or genome-wide analysis of *GRAS* family genes in *Sorghum bicolor* has been published.

**Results:** A total of 81 *SbGRAS* genes were identified based on the *S. bicolor* genome. They were named *SbGRAS01* to *SbGRAS81* and grouped into 13 subfamilies (LISCL, DLT, OS19, SCL4/7, PAT1, SHR, SCL3, HAM-1, SCR, DELLA, HAM-2, LAS and OS4). *SbGRAS* genes are not evenly distributed on the chromosomes, and we found tandem duplication events and segmental duplications of *SbGRAS* genes on *S. bicolor* chromosomes. According to the results of the gene and motif composition, *SbGRAS* members located in the same group contained analogous intron/exon and motif organizations. By quantitative (q) RT-PCR, we quantified the expression of *SbGRAS* members in different plant tissues and in plants exposed to six abiotic stresses at the seedling stage. We also measured gibberellin content under the different abiotic stresses.

**Conclusions:** We identified 81 *SbGRAS* genes and further analyzed their structural composition, and evolution and expression patterns of *SbGRAS* proteins. The latter analysis indicated that *SbGRAS* is important in the course of plant development and its response to abiotic stress. Collectively, this systematic analysis lays the foundation for further study of the functional characteristics of *GRAS* genes of *S. bicolor*.

**Keywords:** *Sorghum bicolor*, *GRAS* gene family, genome-wide analysis, abiotic stress

## Background

Transcription factor (TFs) are a class of proteins that can bind to specific DNA sequences and control the rate of DNA transcription to messenger RNA [1]. This process occurs throughout plant development and regulates the complex gene network in the organism, thereby regulating the basic aspects of biological functions, including cell differentiation, tissue development, organ construction, metabolic synthesis and environmental adaptation [2].

GRAS is a very important TF family that is unique to plants, named after its three members: GAI (GIBBERELIC ACID INSENSITIVE)[3], RGA (REPRESSOR OF GA1-3 MUTANT)[4], and

SCR (SCARECROW)[5]. In general, the coding sequence (CDS) of *GRAS* TFs is 1200–2100 bp in length, and *GRAS* proteins are usually between 400 and 700 amino acids long; however, length and sequence are highly variable [6-9]. Nevertheless, conservation of the *GRAS* family is reflected in the five highly conserved domains at the C terminus of the protein structure: LHR I (leucine-heptad repeat I), VHIID (Val-His-Ile-Ile-Asp), LHR II, PFYRE (Pro-Phe-Tyr-Arg-Glu), and SAW (Ser-Ala-Trp) [9-10]. VHIID is considered the core region and it is highly conserved [8-10]; it is located between the two leucine-rich regions LHR I and LHR II, and can combine with them to form an LHR I–VHIID–LHR II complex which plays an important role in DNA and protein binding [11]. It is worth noting that the two leucine-rich regions do not have the 7 repeated leucine residues that form a leucine zipper [12-14]. The LHR I region has a putative nuclear localization signal near the C terminus, which has been confirmed in the DELLA protein and is similar to the amorphous SV40 (monkey virus) [15]. The latter part of the LHR II domain contains the structure LXXLL (Leu-X-X-Leu-Leu; X represents any amino acid), and it is conserved in over half of the *GRAS* proteins. A similar model of LXXLL can relieve the co-activation complex of mammalian steroid receptors and convert them to homologous nuclear receptors [16]. The LHR I–VHIID–LHR II domain has been confirmed to be involved in the binding of proteins to nucleic acids and other proteins [17-19]. Although PFYRE is not as strictly conserved as the VHIID region, it still exhibits high similarity in all proteins. Aside from Pro, Phe, Tyr, Arg and Glu, there are also Asp (D) and Glu residues in almost all of the FY domains of *GRAS* proteins [20]. The SAW region is near the C terminus and usually contains three sequence units: Trp-X7-Gly (WX7G; X7 represents any 7 amino acids), Leu-Trp (LW) and SAW. The SAW unit is present in almost all *GRAS* proteins. Although the functions of the PFYRE and SAW regions are not known, their high conservation indicates that they are closely associated with *GRAS* protein functions [20]. The *GRAS* proteins also contain a consensus sequence that is similar to SH2, which is a domain in STAT (signal transduction and transcriptional activation) [11]. The conservation of PFYRE and SAW in *GRAS* proteins may indicate an important role in maintaining the protein's structure. At the same time, the N terminus of the *GRAS* protein contains intrinsically disordered regions (IDRs) made up of intrinsically disordered proteins (IDPs), which in turn constitute a unique stretched out structural region that is specific to each *GRAS* member, and affects subfunction differentiation [7]. For example, the DELLA subfamily proteins are well known because they contain the DELLA domain at the N terminus. This differentiation at

the N terminus results in diversification of GRAS proteins [3]. The GRAS gene family contains many subfamilies: their protein sequences have great similarities, but also many differences. Early phylogenetic analysis in *Arabidopsis thaliana* divided GRAS proteins into eight subfamilies: DELLA, LS (LATERAL SUPPRESSOR), SCR, SHR (SHORT ROOT), PAT1 (PHYTOCHROME A SIGNAL TRANSDUCTION), HAM (HAIRY MERISTEM), SCL9 (LISCL; *Lilium longiflorum* SCR-like), and SCL4/7 [21]. Later, Cenci and Rouard [9] proposed that the *GRAS* family members in angiosperms include these eight subfamilies, but also NSP 1, NSP2, DLT (DWARF AND LOW TILLERING), and other subfamilies. Their names are based on the more representative genes in the subfamily, such as DELLA, HAM, DLT, LS, LISCL, PAT1, SCR, SCL3, SHR, and others [6-8].

GRAS proteins are not only differentiated by their structure, but they also have a variety of functions in the biochemical and physiological processes of plants. *SCR* was the first discovered member of the *GRAS* gene family. It is expressed in roots, leaves and vascular bundle sheath cells with SHR. The *SHR* gene regulates SCR-1 (*SCR*) and SCR-2 (*SCL23*), which are involved in the growth of vascular bundle sheath and mesophyll cells in *Arabidopsis* [22]. Both SCR and SHR proteins are involved in regulating the radial growth of *Arabidopsis* roots. As positive regulation factors, they regulate different physiological processes in the formation of radial meristems in roots. The exodermis and endodermis layers are the result of cell division, and *SCR*-knockout plants can only form single-cell tissues, which cannot differentiate into root cortex and endothelial tissues [5]. DELLA proteins not only act as receptors in the gibberellin (GA) reaction [23], but also integrate the signaling pathways of jasmonic acid, auxin, brassinosteroids, and ethylene, constituting a main component of the signaling [24-25]. The N-terminal IDR of the DELLA subfamily and their protein-binding characteristics extend to all plant-specific GRAS proteins, indicating that the N terminus shifts from disordered to ordered. The transformation is related to the specific binding of GRAS proteins [26]. For example, the protein SLN1 (DELLA) of barley was related to the phosphorylation and dephosphorylation of GA signal-related proteins, initiated by the instability of DELLA inhibitors [27]. The AtPAT 1 subfamily is mainly related to the phytochrome-signaling pathway. For example, PAT1-1 (PAT1), PAT1-2 (SCL21) and PAT1-4 (SCL13) have been found in *Arabidopsis* to be located in the phytochrome-signaling pathway downstream of the transduction pathway [38-30]. Overexpression of *Vitis amurensis* PAT1-1 in *Arabidopsis* can enhance its salt and drought tolerance [31]. LISCL was found in *Lilium longiflorum* to induce meiosis-related promoter *lim10*

181 during the microsporogenesis of anthers. LISCL6 (SCL14) was found to play a very important role  
182 in activating stress-inducible promoters, especially salicylic acid (SA)- and 2,4-D-inducible  
183 promoters, thus participating in the heterologous biochemical processes of detoxifying various  
184 substances and harmful endogenous metabolites by regulating its target genes, thereby enhancing  
185 plant tolerance to such harmful substances [37]. *SCL* family genes are programmed, during the  
186 vegetative growth period of plants, to control the formation of side branches, and they are also  
187 present in axillary meristems. For example, overexpression of *Populus euphratica* *SCL7* in *A.*  
188 *thaliana* enhanced tolerance of the latter to salt and drought stress [33-34]. *HAMII-3* (*SCL6*),  
189 *HAMII-2* (*SCL22*) and *HAMII-1* (*SCL27*) in the HAM subfamily inhibit expression of the  
190 protochlorophyllide oxidoreductase C (PORC) gene in light-grown plants, negatively regulating  
191 plant chlorophyll biosynthesis [35]. *SCL3* has been identified as a target gene of DELLA proteins  
192 in *A. thaliana* seedlings [36], and acts as an antagonist of DELLA proteins in controlling the growth  
193 of plants by the regulated GA pathway [36,37]. *NSP1* and *NSP2* are involved in the synthesis of  
194 strigolactones in *Medicago truncatula* and *Oryza sativa*. Strigolactones regulate root branching and  
195 attract arbuscular mycorrhizal fungi, and thus have a very important role [38]. DLT protein has been  
196 found to reduce grain size in rice [39], and it also has a positive regulatory role in the brassinolide-  
197 signaling pathway [40]. Some proteins in the GRAS family play a role through polymerization; for  
198 example, the rice DELLA protein SLENDER RICE1 is capable of homodimerization [41]. Several  
199 GRAS proteins from different subfamilies can undergo heterodimerization, such as SHR–SCR, *M.*  
200 *truncatula* (Mt) NSP1–MtNSP2, MtNSP2–MtRAM1, SCL27–DELLAs, AtSCL3–DELLAs and  
201 PAT1–SCL21, which is usually very important for their functions [18,39-41].

202 Sorghum (*Sorghum bicolor*) is the fifth largest cereal crop in the world. It is an important cash  
203 crop, energy crop and fodder crop, especially in arid and semiarid regions. Sorghum is a C4 plant  
204 with high light efficiency, easy cultivation, strong adaptability, high nutritional value and good  
205 palatability, which exhibits drought resistance, salt–alkali tolerance and other stress-resistance  
206 characteristics. Since the sequencing of the sorghum genome [42], a large number of genetic studies  
207 have been carried out. GRAS proteins have existed throughout the evolutionary history of the entire  
208 plant kingdom. The expansion of this family is closely related to the evolution and diversity of plants  
209 [4-6]. Considering that GRAS is a family of TFs that are unique to higher plants, it plays a vital role  
210 in their growth and development, and especially in the morphogenesis of plant roots, hormone

signaling, light signaling, and plant stress. Most of these factors are characteristic of physiological processes occurring in response to a terrestrial environment. Therefore, the evolution of the *GRAS* gene family provides clues for understanding the adaptive evolution of some C4 plants to environmental changes. In general, the number of members in a particular gene family is relatively similar in different species, but the *GRAS* gene family is different. For example, the number of *GRAS* genes in rice is about twice that in *Arabidopsis* [4-6]. Due to the large differences in the functions of the various subfamilies, the members of the *GRAS* family play important roles in various stages and aspects of plant growth and development [43-46]. In particular, whole-genome analysis of the *GRAS* gene family in different species will help understand the biological functions and evolutionary origins of the *GRAS* genes. The *GRAS* gene family has been extensively studied in many plant species: once the *GRAS* genes of the model organisms *Arabidopsis* and rice were identified [6-8], these genes could be more widely explored in many other species. This family has been identified and analyzed at the whole-genome level in *Solanum lycopersicum* [47], *Vitis vinifera* [48], castor bean [49], *Malus domestica* [50], *Zea mays* [51], *Camellia sinensis* [52], *Gossypium hirsutum* [53], *Capsicum annuum* [54], *Dendrobium catenatum* [55], *Juglans regia* [56], *Fagopyrum tataricum* [57], *Brassica napus* [58], *Citrus sinensis* [59], *Hordeum vulgare* [60], *Manihot esculenta* [61], and others. However, knowledge of the *GRAS* gene family in *S. bicolor* is still very limited. The main gene families that have been identified in the sorghum genome are MADS-box [62], Dof [63], CBL [64], ERF [65], SBP-box [66], HSP [67], LEA [68], and NAC [69], among others. Because the *GRAS* genes play important roles in a variety of physiological processes, it is of great significance to systematically study this family in *S. bicolor*. In the present study, we identified 81 *GRAS* genes and divided them into 13 main groups. Their exon–intron structure, motif composition, gene duplication, chromosome distribution and phylogeny were analyzed. The expression of *GRAS* family members in *S. bicolor* during different biological processes and under abiotic stress was also analyzed. These data provide useful information for the study of the evolutionary relationship and biological function of the *SbGRAS* gene family.

## Results

### Identification of *GRAS* genes in *S. bicolor*

In this study, we used two BLAST methods to identify all possible *GRAS* members in the *S.*

*bicolor* genome. Ultimately, 81 *SbGRAS* genes were identified (Additional file 1: Table S1). They were named *SbGRAS01* to *SbGRAS81* according to their chromosomal location. The basic characteristics were analyzed, including CDS length, protein molecular mass, isoelectric point (pI), domain information and subcellular localization (<http://cello.life.nctu.edu.tw/>) (Additional file 1: Table S1). Of the 81 *SbGRAS* proteins, *SbGRAS80* was the smallest with 174 amino acids, and the largest was *SbGRAS35* with 968 amino acids. The molecular masses of the proteins ranged from 18.99 kDa (*SbGRAS80*) to 107.47 kDa (*SbGRAS35*), and the pI ranged from 4.82 (*SbGRAS66*) to 9.05 (*SbGRAS30*), with a mean of 6.01. The CDS lengths of the *SbGRAS* genes varied greatly, from 522 bp (*SbGRAS80*) to 2904 bp (*SbGRAS35*). The predicted subcellular localization results showed that 37 *SbGRAS* proteins were located in the nuclear region, 24 in the cytoplasm, and 20 in the chloroplast (Additional file 1: Table S1). The number of GRAS TFs in *S. bicolor* exceeded that in *A. thaliana* (32) and rice (57) [8], *Cucumis sativus* (37) [70], *Vitis vinifera* (52) [48] and Tartary buckwheat (47) [57], whereas there were fewer TFs than that in *Malus x domestica* (127) [50] and *Populus trichocarpa* (102) [71]. The ratio of *SbGRAS* genes to total number of genes in the *S. bicolor* genome was about 0.27%, which is more than in *Arabidopsis* (0.11%) [6], rice (0.15%) [8], tomato (0.15%) [8], *Cucumis sativus* (0.14%) [70] and Tartary buckwheat (0.14%) [57], but less than in *Carica papaya* (0.31%) [71] and *Medicago truncatula* (0.29%) [72]. Previous studies have shown that the number and density of GRAS proteins are related to genome size and repetitive events. Therefore, some important GRAS proteins are retained during genome replication to adapt to complex environments [71].

## Multiple sequence alignment, phylogenetic analysis, and classification of *SbGRAS*s

We constructed a phylogenetic tree using the neighbor-joining (NJ) method with a bootstrap value of 1000 based on the amino acid sequences of the 81 identified *SbGRAS*, 33 *AtGRAS* and 50 *OsGRAS* proteins (Figure 1, Additional files 1 and 2: Tables S1 and S2). According to the topological structure of the tree and the classification method proposed by Cenci and Rouard [9], the 164 GRAS proteins in the phylogenetic tree were divided into 13 clades (groups 1–13), consistent with the tree topology and previous classification of the GRAS taxonomic group in angiosperms [9], and indicating no loss of those proteins during *S. bicolor*'s evolution. These findings revealed that GRAS proteins within the reported subfamilies that are present in different

plant species play a basic role in plant development and evolution, similar to those recently reported in previous studies on some other plant species, including *Amborella trichopoda*, *Phoenix dactylifera*, *Vitis vinifera*, *Musa acuminata*, *O. sativa*, *A. thaliana*, *Theobroma cacao* and *Coffea canephora* [9]. Among the 13 subfamilies, LISCL had the most members (39 SbGRASs), and DLT (SbGRAS78), OS19 (SbGRAS58) and SCL4/7 (SbGRAS02) had the fewest (1 *SbGRAS*). There were 7, 7, 6, 5, 4, 3, 3, 2 and 2 *SbGRAS* genes in the PAT1, SHR, SCL3, HAM-1, SCR, DELLA, HAM-2, LAS and OS4 groups, respectively (Figure 1, Additional file 1: Table S1). The phylogenetic tree with *A. thaliana* and *O. sativa* showed that some SbGRASs were tightly grouped with the OsGRASs (bootstrap support  $\geq 70$ ). These results indicate that the GRAS proteins may have evolved further after the separation of monocotyledons and dicotyledons in angiosperms. These proteins may be orthologous to the OsGRASs and have similar functions.

The GRAS proteins of *Arabidopsis* and rice were randomly selected and their LHR I, VHIID, LHR II, PFYRE, and SAW domains were further compared. As shown in Figure 2, the VHIID domain contains a characteristic amino acid sequence, and is considered to be the core region. Although its structure in the different species was highly similar and easy to identify, it was not absolutely conserved. The His and Asp residues in the domain were more conserved. It is worth noting that we divided HAM into groups HAM-1 and HAM-2 based on their genetic and developmental relationships and the results of multiple sequence alignments. Compared to HAM-1, HAM-2 has a more conserved N terminus and VHIID region; it had a smaller number of amino acids in SbGRAS13, SbGRAS35, and SbGRAS66. Among the 81 identified SbGRAS proteins, we observed some domain-loss events in SbGRAS23, SbGRAS26, SbGRAS58, and SbGRAS62, a phenomenon that often occurs in monocots [73]. The N terminus of the SbGRAS proteins contains a highly disordered region, but it shows certain similarities in different subfamilies. For example, the DELLA subfamily protein contains the DELLA domain at its N terminus. This leads to the diversification of GRAS proteins and affects their functional differentiation. Some of the residues in the PFYRE and SAW domains were completely conserved in the different subfamilies, and may therefore be necessary for GRAS protein function.

### Conserved motifs and gene structure analysis of *SbGRAS* genes

To understand the structural components of the *SbGRAS* genes, their exon and intron structures

were obtained by comparing the corresponding genomic DNA sequences ([Figure 3, Additional files 1 and 3: Tables S1 and S3](#)). By comparing the number and position of the exons and introns, we found that the 81 identified *SbGRAS* genes have different numbers of exons, varying from 1 to 5 ([Figure 3A/3B](#)). The 81 *SbGRAS* genes all contained the GRAS domain, and most of the *SbGRAS* genes (54, ~67.5%) contained no introns; 19 *SbGRAS* genes contained 1 intron; *SbGRAS25*, *SbGRAS36*, *SbGRAS38* and *SbGRAS75* contained 2 introns; *SbGRAS52* and *SbGRAS26* contained 3 introns. *SbGRAS51* had the most introns: 4. The 54 intron-less genes were distributed across the other 12 subfamilies, except for the DLT subfamily, and mainly in the LISCL subfamily. Genes without introns are not separated and can therefore be continuously encoded into proteins. Introns are conducive to species evolution: they can increase gene length, increase the frequency of recombination between genes, and they have regulatory effects [74]. Although intron-less genes have no advantage in species evolution or recombination, they tend to respond quickly to stress. Intron-less genes can delay regulatory responses and rapidly regulate the growth and development process [75-76]. Therefore, the *GRAS* gene family may be able to respond quickly to environmental changes. In general, members of the same subfamily had similar gene structures. Members of the groups DELLA, OS19, SCL4/7, DLT, HAM-1, HAM-2, LAS, SCL3 and SCR contained 0 or 1 intron. Further analyses indicated that the LISCL group was most diverse in terms of number of introns.

To further study the characteristic regions of the *SbGRAS* proteins, their motifs were analyzed using an online MEME. A total of 10 distinct conserved motifs (named motifs 1–10) were found ([Figure 3C, Additional file 3: Table S3](#)). As exhibited in Figure 3C, motif 10 was widely distributed in the *SbGRAS*s, except for *SbGRAS23*, *SbGRAS57* and *SbGRAS80*, and was always close to motifs 5 and 3. *SbGRAS* members of the same group usually shared a similar motif composition. For example, group SCL3 contained motifs 10, 6, 7, 2; group PAT1 contained motifs 5, 10, 9, 7, 2, 1, 4; group SHR contained motifs 5, 10, 7, 1. Some motifs were only distributed in specific locations of the pattern. For example, motifs 5 and 10 were always distributed at the start of the pattern, and motif 4 was almost always at the end of the pattern. The functions of most of these conserved motifs remain to be elucidated.

### **Chromosomal spread and gene duplication in *SbGRAS* genes**

A map of the physical position of the *SbGRAS* genes was created based on the most recent *S. bicolor* genome database (Figure 4, Additional files S1 and S4: Tables S1 and S4). Each *SbGRAS* was named according to its physical position from the top to the bottom of *S. bicolor* chromosomes (Chr) 1 to 10. The distribution of the 81 *SbGRAS* genes on the chromosomes was uneven. Interestingly, *SbGRAS* genes were not found on Chr7. We speculate that this is due to fragment loss or chromosome shift during evolution. Chr5 contained the largest number of *SbGRAS* genes (25 genes, ~30.86%), followed by Chr1 (14, ~17.28%); Chr4 and Chr10 contained the least *SbGRAS* genes (4 each, ~4.94%). Chr9 and Chr8 contained 9 (~ 11.11%) and 8 (~ 9.88%) *SbGRAS* genes, respectively. Chr3, Chr6, and Chr8 contained 6 (~ 7.41%), 5 (~ 6.17%), and 6 (~ 7.41%) *SbGRAS* genes, respectively. Interestingly, most members of the *SbGRAS* genes were distributed at the anterior and posterior ends of these 9 chromosomes in *S. bicolor*. In addition, we revealed a large number of genomic duplication events for the *SbGRAS* genes. A chromosomal region within 200 kb containing 2 or more genes is defined as a tandem duplication event [77]. On Chr1, 5, 8 and 9, we found 16 tandem duplication events involving 25 *SbGRAS* genes (Figure 4). *SbGRAS38*, *SbGRAS39*, *SbGRAS40*, *SbGRAS43*, *SbGRAS44*, *SbGRAS48* and *SbGRAS71* each had 2 tandem repeat events (*SbGRAS38* and *SbGRAS37/SbGRAS39*; *SbGRAS39* and *SbGRAS38/SbGRAS40*; *SbGRAS40* and *SbGRAS39/SbGRAS41*; *SbGRAS43* and *SbGRAS42/SbGRAS44*; *SbGRAS44* and *SbGRAS43/SbGRAS45*; *SbGRAS48* and *SbGRAS47/SbGRAS49*; *SbGRAS71* and *SbGRAS70/SbGRAS72*). All of the genes that showed tandem repeat events were members of the same subfamily. It was further discovered that 14 of the 16 pairs were from subfamily LISCL, indicating its important evolutionary role in gene expansion, and indeed, it was the largest subfamily. Only *SbGRAS63/SbGRAS64* and *SbGRAS72/SbGRAS73* were from subfamily SCL3.

In addition, there were 14 pairs of segmental duplications in the *SbGRAS* genes (Figure 5, Additional file 5: Table S5). As shown in Fig. 5, 24 (~29.63%) paralogs were identified in the *SbGRAS* gene family, indicating an evolutionary relationship for these *GRAS* members. The *SbGRAS* genes were unevenly distributed in 10 *S. bicolor* linkage groups (LG). Some linkage groups had more *SbGRAS* genes than others, for example, LG1, LG5, LG8 and LG9 had 4 *SbGRAS* genes, whereas LG2, LG6 and LG10 had only 1 *SbGRAS* gene. Further analysis of these genes' subfamilies showed that all of them were linked within their subfamily. For example, *SbGRAS33* and *SbGRAS65/76* were segmental duplications and they clustered together (subfamily SCR) (Figure 5,

Additional file 5: Table S5). Out of all identified *SbGRAS* genes, group LISCL had the largest number of linked genes (7/24, ~29.17%). In addition, group PAT1 had 4 segmental duplications, while groups HAM-1, HAM-2, LAS, SCL3, SHR has only one pair of segmental duplications (Additional file 5: Table S5). These results suggest that some *SbGRAS* genes may have been produced by gene-replication events—events which have been confirmed to play a major role in the occurrence of new functions in *S. bicolor* evolution and amplification of the *SbGRAS* gene family.

### Syntenic analysis of *SbGRAS* genes

To further infer the phylogenetic mechanisms in the *SbGRAS* family, we constructed six comparative syntenic maps of *S. bicolor* with six representative species: three dicotyledons (*A. thaliana*, *C. annuum* and *Solanum lycopersicum*) and three monocotyledons (*O. sativa*, *Brachypodium distachyon* and *Z. mays*) (Figure 6, Additional file 6: Table S6). A total of 51 *SbGRAS* genes showed syntenic relationships with those in *Arabidopsis* (6), followed by *C. annuum* (9), tomato (14), *B. distachyon* (37), indica rice (39) and maize (65) (Additional file 6: Table S6). The number of orthologous pairs between the other six species (*Arabidopsis*, *C. annuum*, tomato, *B. distachyon*, indica rice and maize) was 14, 15, 27, 51, 56 and 90, respectively.

Some *SbGRAS* genes were found to be associated with at least one syntenic gene pair among the six plants (especially between *S. bicolor* and *Z. mays* *GRAS* genes), such as *SbGRAS01*, *SbGRAS12*, *SbGRAS24*, *SbGRAS31*, suggesting that these orthologous pairs already existed before the ancestral divergence, and thus indicating that these genes may have played an important role in the *GRAS* gene family during evolution. Interestingly, some collinear gene pairs (with 12 *SbGRAS* genes) identified between *S. bicolor* and *B. distachyon*/indica rice/maize were not found between *S. bicolor* and *Arabidopsis*/*C. annuum*/tomato, such as *SbGRAS02*, *SbGRAS03*, *SbGRAS18*, *SbGRAS23*, *SbGRAS29*, *SbGRAS30*, *SbGRAS34*, *SbGRAS35*, *SbGRAS36* and *SbGRAS42*. This suggests that these orthologous pairs formed after the divergence of dicotyledonous and monocotyledonous plants (Additional file 6: Table S6). Significantly, some collinear *GRAS* gene pairs identified between *S. bicolor* and indica rice/maize/*B. distachyon* were anchored to highly conserved syntenic blocks that spanned 18 genes. In contrast, those between *S. bicolor* and *Arabidopsis*/*C. annuum*/tomato were all located in syntenic blocks that had less than 10 orthologous gene pairs. This might be related to the phylogenetic relationship between *S. bicolor* and the other six plant species.

To better understand the evolutionary constraints acting on the *SbGRAS* gene family, the *SbGRAS*s were subjected to Tajima's D Neutrality Test [78-79]. Calculations gave  $D = 7.25577$ ; the large deviation from 0 suggests that the *SbGRAS* gene family might have experienced strong purifying selective pressure during evolution ([Additional file 7: Table S7](#)).

### **Evolutionary analysis of the *SbGRAS* genes and *GRAS* genes of several different species**

To analyze the evolutionary relations between the trihelix family of *SbGRAS* proteins among *S. bicolor* and six plants (*A. thaliana*, *C. annuum*, *Solanum lycopersicum*, *B. distachyon*, *O. sativa* subsp. *indica*, *Z. mays*), an unrooted NJ tree with 10 conserved motifs according to the MEME web server was constructed using the NJ method of Geneious R11 according to the protein sequences of the 81 identified *SbGRAS* genes and the six other plants' trihelix genes ([Figure 7, Additional file 3: Table S3](#)). The distribution of *SbGRAS*s in the phylogenetic tree was relatively widely dispersed. As shown in Figure 7, the *SbGRAS* proteins tended to gather with the *GRAS* proteins of *O. sativa* and *Z. mays*, suggesting that they are more closely related. Most of the *GRAS* proteins from the six studied plants contained motifs 5, 10 and 2. In addition, several motifs were only present in the *GRAS* proteins of a few specific *SbGRAS* branches, such as motif 1. Motif 5 was distributed between motifs 10 and 6, and motif 7 was distributed between motifs 2 and 8. Motif 10 was always distributed at the start of the pattern and motif 4 was almost always distributed at the end of the pattern. We also found that the *GRAS* proteins of *O. sativa*, *Z. mays* and *S. bicolor* on the same branch generally had similar motif compositions, and similar serial motifs tended to cluster in specific *GRAS* protein subfamilies, indicating potential functional similarities between the *GRAS* proteins.

### **Expression patterns of *SbGRAS*s in several organs**

To further analyze the potential roles of *SbGRAS* genes, we randomly selected 1 gene in each subfamily and tested the expression of these 13 representative genes in six organs (anthers, styles, fruit, roots, stems, leaves) by qRT-PCR assay ([Figure 8A](#)). The expression patterns of the *SbGRAS* genes changed substantially among the six plant organs, with some exhibiting preferential expression across the detected tissues of *S. bicolor*. Most of genes were expressed in all organs; 2 genes (*SbGRAS11* and *SbGRAS78*) showed the highest expression level in the anther; *SbGRAS28* and *SbGRAS29* showed the highest expression level in the style, and the highest expression of

*SbGRAS02*, *SbGRAS14* and *SbGRAS27* was found in the fruit; 2 genes (*SbGRAS03* and *SbGRAS31*) were highly expressed in the stem. The results showed diverse transcriptional abundance of *SbGRAS* genes in different tissues and organs, indicating that these genes have multiple functions in sorghum growth and development of. For example, DELLA protein is the main negative regulator of GA-signal transduction [80]; the expression level of *SbGRAS03* of the DELLA family was high during the fruit-filling stage, suggesting an important role in sorghum fruit development. Correlations of *SbGRAS* gene expression among the six organs were also studied (Figure 8B). We found that the expression of different genes in the plant organs was significantly correlated, indicating their possible synergistic role. Most *SbGRAS* genes showed significant positive correlations; for example, *SbGRAS04*, *SbGRAS11*, *SbGRAS28*, *SbGRAS29*, and *SbGRAS78* were all highly expressed in the style and anther, and their expression was significantly positively correlated. However, these same genes were significantly negatively correlated with *SbGRAS02*, *SbGRAS03*, and *SbGRAS31*.

#### **Expression patterns of *SbGRAS* genes in response to different abiotic stresses**

To further determine whether the expression of *SbGRAS* genes is influenced by different abiotic stresses, 13 *SbGRAS* members were examined for their expression under six abiotic stresses: strong ultraviolet radiation (UV), flooding, polyethylene glycol (PEG), NaCl, heat and cold treatments. We ran qRT-PCR experiments to analyze the 13 *SbGRAS* members' expression patterns in roots, leaves and stems in response to the different treatments (Figure 9A). Some *SbGRAS* genes were significantly induced/repressed by a number of the abiotic stress treatments. Expression of most of these genes was significantly altered in the early stage of the treatment (Figure 9A). Among them, some *SbGRAS*s showed changes in expression that were similar or opposite at different times and in different organs. For example, under UV, flooding, PEG, and NaCl treatment, the expression level of *SbGRAS04* decreased significantly in the roots, stems and leaves, indicating its rapid inhibition by these stresses. But its expression was totally opposite under heat and cold stresses: it was initially significantly upregulated and then downregulated in roots, stems and leaves, which expression pattern remained constant. Under all treatments, the expression of *SbGRAS03* (DELLA) in the roots increased significantly, whereas in the stems and leaves, it was initially significantly upregulated and then downregulated. The expression level of *SbGRAS58* increased significantly under UV and flooding treatments, but decreased gradually under PEG treatment. We also looked at the

correlations between the *SbGRAS* genes' expression (Figure 9B). There were negative correlations among most *SbGRAS* genes. However, the expression levels of *SbGRAS11*, *SbGRAS27*, *SbGRAS28*, *SbGRAS31* and *SbGRAS33* still showed a significant positive correlation ( $p < 0.05$ ), indicating that they may belong to a positive regulatory network (Figure 9B). We further studied the content of GAs in these tissues under the different stresses (Figure 9C). Under flooding, PEG, NaCl, heat and cold treatments, GA content gradually decreased in roots, stems, and leaves. Under UV treatment, root GA content decreased significantly, but there was no change in the stems or leaves. This might be due to the reduction in GA inhibition level induced by UV irradiation.

## Discussion

The *GRAS* gene family plays an important role in regulating plant growth and development, and in coping with environmental stress [8]. Although the whole genome sequence of sorghum has been published, there are few research reports on the *GRAS* gene family in this plant. This study systematically analyzed the *GRAS* gene family in *S. bicolor*, and identified a total of 81 *SbGRAS* genes. All *SbGRAS* proteins showed significant differences in structure, indicating a high degree of complexity. The CDS length of the *SbGRAS* TFs had a wide range of 522 (*SbGRAS80*) to 2904 (*SbGRAS35*) bp, and the lengths of the *GRAS* proteins varied between 174 and 968 amino acids, indicating great variability [6-8]. This variation may be related to gene-duplication events or genome size [48]. Our phylogenetic analysis divided *SbGRAS* genes into 13 subfamilies: *LISCL*, *SCL4/7*, *DELLA*, *HAM-1*, *SHR*, *HAM-2*, *PAT1*, *SCR*, *OS4*, *SCL3*, *LAS*, *OS19* and *DLT* (Figure 1). According to the constructed phylogenetic tree, we identified at least one *SbGRAS* protein in each subgroup of *A. thaliana*, indicating that the divergence of the *GRAS* family may have occurred earlier than that of monocotyledons and dicotyledons, with some new subfamily members being produced as evolution proceeded. Although there has been some fine-tuning in different plant species, the previous phylogenetic clustering of this family is basically the same, indicating that the *GRAS* genes in angiosperms are highly diverse [6-9]. Among the 13 subfamilies, *LISCL* had the most members (39, ~48.15%), while *DLT* (*SbGRAS78*), *OS19* (*SbGRAS58*), and *SCL4/7* (*SbGRAS02*) had the least (1 member). This is similar to other plants, such as *A. thaliana* [6-8], rice [6-8], and maize [51], indicating that these *GRAS* gene families may have had strong partial differentiation capabilities in the long-term evolutionary process. However, there is no research to prove that this differentiation

is beneficial to the differentiation process of herbs and woody plants. Most of these genes share five main conserved domains: LHR I, VHIID, LHR II, PFYRE and SAW (Figure 2). The core VHIID domain contains the characteristic sequence of Val, His, Ile and Asp. The His and Asp residues are more conserved. In previous studies, some important domains have been shown to have important structural functions. For example, the LHR I–VHIID–LHR II complex can effectively mediate protein–protein and protein–DNA interactions [81–82]; a point mutation in the conserved LHR I region of NSP2 interrupted the formation of the NSP1–NSP2 complex, an interfamily heteropolymer important for nodulation signaling, and hence interfered with nodule development in *M. truncatula* [19]; mutations in the LHR II and PFYRE domains in rice SCL7 destroyed its DNA-binding ability [82]. We divided the HAM subfamily into two groups, HAM-1 and HAM-2, based on their genetic and developmental relationship and the multiple-sequence-alignment results. HAM-2 has more conserved N-terminal and VHIID regions than HAM-1; it has fewer amino acids, and it is only found in *SbGRAS13*, *SbGRAS35* and *SbGRAS66*. We also observed some domain-loss events in *SbGRAS23*, *SbGRAS26*, *SbGRAS58* and *SbGRAS62*. Although previous studies on the functions of the PFYRE and SAW domains were not performed in depth, we observed that some residues in these regions are absolutely conserved in different subfamilies, and that these residues may be required for GRAS proteins of different subfamilies to function. In general, an IDR in an IDP allows the proteins to recognize and interact with various partners, which is essential for molecular function. Upon IDP binding, the IDR often transitions from disordered to ordered, allowing the IDP to recognize and bind different partners at various binding interfaces [7, 26, 83]. Here, we found that the *SbGRAS* protein contains a highly variable N-terminal region, which is consistent with the view that its N terminus is essentially disordered, and this explains the functional differentiation of *SbGRAS* protein.

We analyzed the exon and intron structures of the 81 identified *SbGRAS* genes (Figure 2, [Additional file 1: Table S1](#)); the number of exons per gene ranged from 1 to 5 (Figure 2A/2B). The proportion of *SbGRAS* genes without introns was higher (54, ~66.67%) than in plum blossom (82.2%), *Arabidopsis* (67.6%), rice (55%) and poplar (54.7%) [46, 71, 84], indicating that there is a close evolutionary relationship among GRAS members. Intron-less genes have also been found in other large gene families, such as the DEAD box RNA helicase [85] and F-box TF families, and the small auxin-up RNA (Saur) gene family [86]. Although there are no genes with introns in

prokaryotic genomes, one study [51] showed that plant *GRAS* genes originated from the prokaryotic genes mainly by horizontal gene transfer and by duplication events during their evolution. This phenomenon explains the large number of intron-less *GRAS* genes based on sequence homology and classification [87]. Zhang et al. [88] reported a method for extracting *GRAS* genes from a prokaryotic bacterial genome by horizontal gene transfer. Furthermore, the characteristic regions of the SbGRAS proteins were further studied, and their motifs were analyzed by online MEME. Ten different conserved motifs were found, and *SbGRAS* members in the same group usually had a similar motif composition (Figure 3). It is possible that the transcriptional regulators in a taxonomic group have the most recent common evolutionary origin and molecular function, which makes this an effective and practical method to predict the function of unknown proteins [57].

We analyzed the distribution of the 81 *SbGRAS* genes on sorghum Chr1–Chr10 and found it to be uneven. Interestingly, *SbGRAS* was not found on Chr7 of sorghum. We speculate that this is due to fragment loss or chromosome shift during evolution. Gene amplification refers to the duplication of one or more copies of a DNA fragment. This DNA fragment can be a small segment of the genome sequence, the entire chromosome, or even the entire genome. The amplification can occur in a variety of ways, including tandem duplication events and segmental duplication events. Gene amplification is a very important driving force in genome evolution, which can lead to the emergence of new functional genes and the differentiation of new species, thereby rendering plants more adaptable to harsh environments during evolution [71]. Tandem repeat events occur more frequently in plant genomes, such as *Arabidopsis* and rice, affecting about 10% of the genes [6]. Tandem replication only amplifies the number of one type of gene, so that a gene that was originally a copy in the ancestral genome forms a gene cluster composed of multiple genes after tandem replication; this plays an extremely important role in the expansion of plant gene families. Analysis of these genes shows that tandem replication tends to be amplified, which can improve plant tolerance to abiotic stress. Compared to *Arabidopsis* and rice, there are more GRAS proteins in sorghum, indicating that there may have been more gene duplication events in sorghum or a higher frequency of retained copies after replication. We found 16 tandem repeat events in the SbGRAS proteins, involving 25 genes (~30.9%), which is higher than in *Arabidopsis* 2/34 [6], plum 10/45 [46], tomato 15/53[47] or rice 17/60 [6], but lower than in poplar 40/106 [87]. It is worth noting that the two narrow regions of Chr5 contain 4 pairs and 7 pairs, respectively, of closely homologous

genes (Figure 4, belonging to the LISCL subfamily). We found that all *SbGRAS* genes that had tandem repeats came from the same subfamily, and mainly occurred in the LISCL subfamily (14, ~87.5%). This implies that the retention of gene copies is biased to a certain extent after the whole genome is replicated, while the structure does not produce great differences, and the retention and loss of different subgenomes are also different. In addition, some studies have found that if there is an interaction between the protein and other products encoded by the gene, this type of gene will be biased after the replication event has occurred [89]. In addition, the *SbGRAS* genes had 14 pairs of fragments. Duplication events were found (Figure 4, Additional file 5: Table S5) in 23 *SbGRAS* genes (~28.8%). The contribution of tandem repeats (25 *SbGRAS* genes) to the increase in sorghum *GRAS* members was slightly greater than that of fragment repeats (23 *SbGRAS* genes). Further analysis of these gene subfamilies showed that they are all linked within subfamilies. These results indicate that some *SbGRAS* genes may have been generated by gene-duplication events, further confirming that duplication events are the most common mechanism leading to the rapid expansion of *GRAS* family members in different species.

The analysis of gene-expression profile can provide important clues to understanding potential biological functions. There are many members of the GRAS TF family with diverse functions, but the current research in plants has not been particularly deep and mainly relates to the model plants *Arabidopsis* and rice. The role of GRAS TFs in other species remains to be explored. Roots, stems, leaves, flowers and fruit are the main organs in all angiosperms [90]. In this study, the expression of *SbGRAS* genes from 13 subfamilies was studied in the stamens, styles, and main organs of *S. bicolor*. Studies have shown that GRAS TFs play an important role in the development of flowers and fruit in many plants. *SbGRAS02*, *SbGRAS14* and *SbGRAS27* showed the highest expression levels in fruit (Figure 8A). *SbGRAS03* of the DELLA family also demonstrated higher expression during the fruit-filling stage, suggesting an important role in the development of sorghum fruit. However, specific functions need to be analyzed through in-depth experiments. In addition, the correlations of the various genes' expression among six plant organs was also studied. The expression of most *SbGRAS* genes was significantly positively correlated, suggesting synergistic activity. In summary, these results revealed the functions of some GRAS TFs and their self-regulation. We also studied the responses of these 13 typical *SbGRAS* genes to six abiotic stresses in different organs, and found that almost all of them exhibited significant differential expression

under stress (more than 2-fold change). For example, under PEG stress, the expression of 9 *GRAS* genes was upregulated in roots, 11 in stems, and 9 in leaves. Interestingly, some genes had completely opposite expression patterns compared to the other stresses, such as *SbGRAS04*. The correlations indicated that GRAS TFs participate in a complex cross-regulatory network, and that they may respond to various adverse conditions through coordinated or antagonistic mechanisms. Further research is warranted to explore the relationship between these genes. Interestingly, most *SbGRAS* genes showed significant negative regulation on the expression heat map, although we found a significant positive correlation in the expression of 5 genes. If we consider expression patterns and complex protein interactions, then we can speculate that there is a network of feedback mechanisms that coordinate the expression of multiple genes.

DELTA protein is a negative regulator of the GA pathway, and plays an important role in plant growth and disease resistance by regulating a variety of hormone signals. On the one hand, OsSLR1 has a positive regulatory effect on the levels of 2 mitogen-activated protein kinases and 4 WRKY transcripts, as well as on levels of jasmonic acid, ethylene and H<sub>2</sub>O<sub>2</sub>. On the other hand, silencing OsSLR1 can increase the level of defense-related compounds, phenolic acids, lignin and cellulose, and the resistance of rice to brown planthopper [91]. In this study, *SbGRAS03* was significantly expressed in the root, stem, leaves and fruit. At the same time, it was significantly induced under UV, flooding, PEG and NaCl treatments at the seedling stage, and its expression level gradually increased and then fell in the stems and leaves. It reached its highest expression level in 2h. OsGRAS19 and the brassinosteroid signal-transduction pathway can regulate grain size by promoting cell division and regulating the number of epidermal cells on glumes. Overexpression of the *OsGRAS19* gene or CRISP/Cas9 mutants showed an effect on grain size. *SbGRAS58* had positive effects in the roots, stems and leaves [92], and more significant expression in the fruit. What is interesting is that there was also significant expression in the style, suggesting that it may participate in a complex biological development network. Not only that, but *SbGRAS58* exhibited completely different response modes under different stresses. For example, under UV and flooding, expression level in the rhizomes showed a significant gradual upward trend; under heat and PEG stress, the expression level showed a continuous downward trend; under heat stress, the expression level showed a trend of first increasing and then decreasing. We found some differences in the expression patterns of the selected HAM-1 (*SbGRAS31*) and HAM-2 (*SbGRAS13*) genes. We further analyzed

their expression in different tissues. The expression of *SbGRAS13* in the fruit and style was significantly higher than that of *SbGRAS31*, while in the stem and anther, the opposite was seen. HAM-1 and HAM-2 are two different HAM subfamily members, and *SbGRAS13* and *SbGRAS31* show completely inconsistent expression trends. For example, under UV and flooding stress, *SbGRAS13* first increases and then decreases in roots and stems. The expression trend of *SbGRAS31* first decreased and then increased, indicating that more detailed differentiation may lead to great functional differences in the response to environmental stress. *Arabidopsis* SHR may interact with SCR and SCL23 subfamily proteins to form the SHR–SCR–SCL23 module that regulates root endoderm development [93]. The 7 SHRs and 4 SCRs identified in sorghum may share functions with SHR and SCR in *Arabidopsis*, as their functions in rice and *Arabidopsis* have been shown to be conserved [93]. Nevertheless, like the *Arabidopsis* SHR subfamily, some or all of the loss in gene function of the broader SHR and SCR subfamilies in sorghum needs to be carefully studied in the future, because most of their gene members are expressed at low levels (Figures 8A and 9A). DELLA protein regulates plants' stress tolerance [94]. In our qRT-PCR results, *SbGRAS03* had an extremely high expression response under all stress treatments, although the expression levels in different tissues were inconsistent, indicating their important role in coping with adversity. DELLA protein inhibits plant growth and development, and GA promotes plant growth and development by degrading DELLA protein [94]. The degradation of DELLA is initiated by the formation of the GA–GID1–DELLA complex, which is then recognized by the specific ubiquitin E3 ligase complex (SCFSLY1/GID2), which marks the DELLA protein for degradation by the 26S proteasome [94]. DELLA also plays an important role in fruit development and abiotic stresses [95-97]. Therefore, we investigated the changes in GA content in seedling tissues under six kinds of stress; the GA content of roots, stems, and leaves gradually decreased under flooding, PEG, NaCl, cold and heat treatment. Under UV light treatment, root GA content was significantly reduced, but remained unchanged in the stems and leaves. This might be due to the reduced level of GA inhibition caused by UV radiation. In general, *SbGRAS* genes have a potential regulatory role in plant development and responses to stress. These overall findings provide insights into the potential functional roles of sorghum *GRAS* genes, and help understand the developmental process in sorghum toward genetic improvement of environmental resistance.

## Conclusion

In this study, 81 *GRAS* gene family members were identified in the genome of *S. bicolor*, and phylogenetic analysis indicated that these *SbGRAS* genes could be classified into 13 subfamilies. Most of the *SbGRAS* genes were intron-less. It was found that gene-replication events may have produced some *SbGRAS* genes, with tandem duplication contributing more to the expansion of the *SbGRAS* gene family than segmental duplication. Phylogenetic comparison and synteny analysis of *GRAS* genes from six typical plant species provided valuable clues about the evolutionary characteristics of *GRAS* gene family members in *S. bicolor*. Finally, the expression patterns of the *GRAS* members of *S. bicolor* under abiotic stresses and in different tissues or at different developmental stages were explored using qRT-PCR, which indicated the important roles of *SbGRAS* genes in growth and development.

## Methods

### Gene identification

We downloaded the *S. bicolor* whole genome sequence information from the Ensembl Genomes website (<http://ensemblgenomes.org/>). The *GRAS* gene family of *S. bicolor* was searched by two BLASTP methods [98, 99]: first, with BLASTp (score value  $\geq 100$  and e-value  $\leq 1e-10$ ), all possible *GRAS* proteins were identified from the *S. bicolor* genome referring to trihelix protein sequences of *Arabidopsis*. Then, the hidden Markov model (HMM) file corresponding to the *GRAS* trihelix domain was downloaded from the Pfam protein family database (<http://pfam.xfam.org/>). We used both HMMER3.0 (default parameters) with a cutoff of 0.01 (<http://plants.ensembl.org/hmmer/index.html>) [100] and SMART (<http://smart.embl-heidelberg.de/>) [101, 102] to ascertain the presence of the *GRAS* domain, and the HMMER was used to further verify the results. By using the tools from the ExPASy website ([https://web.expasy.org/compute\\_pi/](https://web.expasy.org/compute_pi/)), we obtained the sequence length, molecular weight, pI and subcellular localization of the 81 identified *GRAS* proteins.

### GRAS structure

The *SbGRAS* domain sequences of the characterized *GRAS* proteins were used to create multiple protein sequence alignments using ClustalW with default parameters [103]. The deduced

amino acid sequences in the GRAS domains were then adjusted manually using GeneDoc software. To study the structural differences between the *SbGRAS* genes, conserved motifs were studied in the encoded GRAS proteins [105, 106]. We used the gene structure display server (GSDS: <http://gsds.cbi.pku.edu.cn>) online program to analyze the exon–intron structure of the *SbGRAS* genes based on CDS length and corresponding full-length sequence. An online MEME program (<http://meme.nbcr.net/meme/intro.html>) was used to analyze the protein sequences with the following parameters: optimum motif width of 6 ~ 200, and maximum number of motifs, 10 [99, 106, 107].

### **Chromosomal distribution and gene duplication**

All *SbGRAS* genes were mapped to *S. bicolor* chromosomes based on physical location information from the database of the *S. bicolor* genome using Circos [108]. Analysis of *SbGRAS* gene-replication events was conducted using multiple collinear scanning toolkits (MCScanX) with the default parameters [109]. We analyzed *GRAS* gene homology between *S. bicolor* and six other plants (*A. thaliana*, *C. annuum*, *Solanum lycopersicum*, *B. distachyon*, *O. sativa* subsp. *indica*, *Z. mays*) using Dual Synteny Plotter (<https://github.com/CJ-Chen/TBtools>).

### **Phylogenetic analysis and classification of *SbGRAS* gene family**

According to the classification of the *AtGRAS*s, the identified *SbGRAS* genes were divided into different groups. The phylogenetic trees were inferred using the NJ method of MEGA X via Geneious R11 with the BLOSUM62 cost matrix, the Jukes–Cantor model, global alignment with free end gaps and bootstrap value of 1000. The full-length amino acid sequences of the GRAS proteins (Additional file 1: Table S1) derived from *A. thaliana*, *C. annuum*, *Solanum lycopersicum*, *B. distachyon*, *O. sativa* subsp. *indica*, and *Z. mays* (UniProt<https://www.uniprot.org/>) combined with the newly identified *SbGRAS*s were used for phylogenetic analysis.

### **Plant materials, growth conditions, and abiotic stress in *S. bicolor***

The *S. bicolor* variety Hongyingzi materials used in the experiment were supplied by Prof. Cheng Jianping of Guizhou University. 'Hongyingzi' has been grown in the greenhouse at Guizhou University since 2019. *S. bicolor* plants were grown in pots filled with soil and vermiculite (1:1) in

a growth room with a 16 h/25°C day and 8 h/20°C night regime, and a relative humidity of 75%. We collected the stems, roots, leaves, fruit, anthers and styles separately from five plants with good growth and similar growth conditions, and quickly placed them in liquid nitrogen for storage at –80 °C for further use. Several *SbGRAS* genes were selected to investigate their expression patterns in response to various stresses. *S. bicolor* plants at the seedling stage (21 days) were selected for the abiotic stress treatments, which included salt treatment (5% NaCl), water flooding (whole plant), drought (30% PEG6000), UV radiation (70 μW/cm<sup>2</sup>, 220 V, 30 W), high temperature (40 °C), and low temperature (4 °C). Each stress treatment had five replicates; qPCR detection and hormone (GA) analysis were carried out after sampling at 2 h and 24 h, respectively. The collected samples were stored at –80 °C for subsequent analysis.

#### **Total RNA extraction, cDNA reverse transcription and qRT-PCR analysis**

The cDNA was produced with a 1 mg RNA sample using a PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa) and SYBR Premix Ex Taq II (TaKaRa) [100]. The sequencing was performed in an Illumina GAII sequencer following the manufacturer's instructions [106, 107]. Gene-expression analysis of the selected genes was performed by qRT-PCR, and repeated at least three times, with primers designed by Primer 5.0 (Additional file 8: Table S8). We used the *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) gene, which was stably expressed at each growth stage in almost all tissues, as the internal control [110]. The correlations of expression data were calculated according to the  $2^{-(\Delta\Delta CT)}$  method [111].

#### **Endogenous GA analysis**

We added 20 mL of 80% ethanol to 1 g of fresh sample. Then the samples were extracted three times with ultrasound, for 1 h each time. The extract was concentrated once at low temperature and mixed with water, and then extracted with an equal volume of N-butanol; the N-butanol layer was dried under a nitrogen stream. Then, 20 mg of the dried sample was dissolved in 5 mL methanol (MS grade), and filtered through a 0.22-μm membrane. The chromatographic conditions were: a C18 column (2.1 mm × 75 mm, 2.7 μm) held at a constant 40°C; injection volume 5 μL; the mobile phase was acetonitrile (A) and deionized water (B); the mobile phase gradient elution was 0 min~1 min, 20% A; 1 min~1.5 min, 80% A; 1.5 min~4.5 min, 80% A; 4.5 min~6 min, 20% A; 6 min~8

min, 20% A.

## Statistical analysis

Analysis of variance (ANOVA) was performed with JMP6.0 software (SAS Institute), and compared by least significant difference (LSD) at the 0.05 and 0.01 levels. The histogram was drawn with Origin 8.0 software (SAS Institute).

## References

- [1] Latchman DS. Transcription factors: an overview. *Int J Biochem Cell Biol.* 1997 Dec;29(12):1305-1312.
- [2] Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, Adam L, Pineda O, Ratcliffe OJ, Samaha RR, Creelman R, Pilgrim M, Broun P, Zhang JZ, Ghandehari D, Sherman BK, Yu G. Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. *Science.* 2000 Dec 15;290(5499):2105-2110.
- [3] 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 Dec 1;11(23):3194-3205.
- [4] Silverstone AL, Ciampaglio CN, Sun T. The Arabidopsis RGA gene encodes a transcriptional regulator repressing the gibberellin signal transduction pathway. *Plant Cell.* 1998 Feb;10(2):155-169.
- [5] Di Laurenzio L, Wysocka-Diller 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 Aug 9;86(3):423-433.
- [6] 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 Mar;54(4):519-532.
- [7] Sun X, Xue B, Jones WT, Rikkerink E, Dunker AK, Uversky VN. A functionally required unfoldome from the plant kingdom: intrinsically disordered N-terminal domains of GRAS proteins are involved in molecular recognition during plant development. *Plant Mol Biol.* 2011 Oct;77(3):205-223.

751 [8] Liu X , Widmer A . Genome-wide Comparative Analysis of the GRAS Gene Family in Populus,  
752 Arabidopsis and Rice. *Plant Molecular Biology Reporter*, 2014, 32(6):1129-1145.

753 [9] Cenci A, Rouard M. Evolutionary Analyses of GRAS Transcription Factors in Angiosperms.  
754 *Front Plant Sci*. 2017 Mar 2;8:273.

755 [10] Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN. The GRAS gene family in  
756 Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE  
757 genes. *Plant J*. 1999 Apr;18(1):111-9.

758 [11] Richards DE, Peng J, Harberd NP. Plant GRAS and metazoan STATs: one family? *Bioessays*.  
759 2000 Jun;22(6):573-577.

760 [12] Sun, XL, Xue, et al. A functionally required unfoldome from the plant kingdom: intrinsically  
761 disordered N-terminal domains of GRAS proteins are involved in molecular recognition during  
762 plant development. *PLANT MOL BIOL*, 2011, 2011,77(3)(-):205-223.

763 [13] Dill A, Jung HS, Sun TP. The DELLA motif is essential for gibberellin-induced degradation of  
764 RGA. *Proc Natl Acad Sci U S A*. 2001 Nov 20;98(24):14162-14167.

765 [14] Itoh H, Ueguchi-Tanaka M, Sato Y, Ashikari M, Matsuoka M. The gibberellin signaling  
766 pathway is regulated by the appearance and disappearance of SLENDER RICE1 in nuclei. *Plant*  
767 *Cell*. 2002 Jan;14(1):57-70.

768 [15] Raikhel N. Nuclear targeting in plants. *Plant Physiol*. 1992 Dec;100(4):1627-1632.

769 [16] Heery DM, Kalkhoven E, Hoare S, Parker MG. A signature motif in transcriptional co-  
770 activators mediates binding to nuclear receptors. *Nature*. 1997 Jun 12;387(6634):733-736.

771 [17] De Lucas M, Davière JM, Rodríguez-Falcón M, Pontin M, Iglesias-Pedraz JM, Lorrain S,  
772 Fankhauser C, Blázquez MA, Titarenko E, Prat S. A molecular framework for light and gibberellin  
773 control of cell elongation. *Nature*. 2008 Jan 24;451(7177):480-484.

774 [18] Hou X, Lee LY, Xia K, Yan Y, Yu H. DELLAs modulate jasmonate signaling via competitive  
775 binding to JAZs. *Dev Cell*. 2010 Dec 14;19(6):884-894.

776 [19] Hirsch S, Kim J, Muñoz A, Heckmann AB, Downie JA, Oldroyd GE. GRAS proteins form a  
777 DNA binding complex to induce gene expression during nodulation signaling in *Medicago*  
778 *truncatula*. *Plant Cell*. 2009 Feb;21(2):545-557.

779 [20] Levy DE, Darnell JE Jr. Stats: transcriptional control and biological impact. *Nat Rev Mol Cell*  
780 *Biol*. 2002 Sep;3(9):651-662.

781 [21] Bolle C. The role of GRAS proteins in plant signal transduction and development. *Planta*. 2004  
782 Mar;218(5):683-692.

783 [22] Cui H, Kong D, Liu X, Hao Y. SCARECROW, SCR-LIKE 23 and SHORT-ROOT control  
784 bundle sheath cell fate and function in *Arabidopsis thaliana*. *Plant J*. 2014 Apr;78(2):319-327.

785 [23] Sun TP. The molecular mechanism and evolution of the GA-GID1-DELLA signaling module  
786 in plants. *Curr Biol*. 2011 May 10;21(9):R338-345.

787 [24] Bai MY, Shang JX, Oh E, Fan M, Bai Y, Zentella R, Sun TP, Wang ZY. Brassinosteroid,  
788 gibberellin and phytochrome impinge on a common transcription module in *Arabidopsis*. *Nat Cell*  
789 *Biol*. 2012 Aug;14(8):810-817.

790 [25] Wild M, Davière JM, Cheminant S, Regnault T, Baumberger N, Heintz D, Baltz R, Genschik  
791 P, Achard P. The *Arabidopsis* DELLA RGA-LIKE3 is a direct target of MYC2 and modulates  
792 jasmonate signaling responses. *Plant Cell*. 2012 Aug;24(8):3307-3319.

793 [26] Sun X, Jones WT, Rikkerink EH. GRAS proteins: the versatile roles of intrinsically disordered  
794 proteins in plant signalling. *Biochem J*. 2012 Feb 15;442(1):1-12.

795 [27] Fu X, Richards DE, Ait-Ali T, Hynes LW, Ougham H, Peng J, Harberd NP. Gibberellin-  
796 mediated proteasome-dependent degradation of the barley DELLA protein SLN1 repressor. *Plant*  
797 *Cell*. 2002 Dec;14(12):3191-3200.

798 [28] Bolle C, Koncz C, Chua NH. PAT1, a new member of the GRAS family, is involved in  
799 phytochrome A signal transduction. *Genes Dev*. 2000 May 15;14(10):1269-78. PMID: 10817761;  
800 PMCID: PMC316623.

801 [29] Torres-Galea P, Huang LF, Chua NH, Bolle C. The GRAS protein SCL13 is a positive regulator  
802 of phytochrome-dependent red light signaling, but can also modulate phytochrome A responses.  
803 *Mol Genet Genomics*. 2006 Jul;276(1):13-30.

804 [30] Torres-Galea P, Hirtreiter B, Bolle C. Two GRAS proteins, SCARECROW-LIKE21 and  
805 PHYTOCHROME A SIGNAL TRANSDUCTION1, function cooperatively in phytochrome A  
806 signal transduction. *Plant Physiol*. 2013 Jan;161(1):291-304.

807 [31] Yuan Y, Fang L, Karungo SK, Zhang L, Gao Y, Li S, Xin H. Overexpression of VaPAT1, a  
808 GRAS transcription factor from *Vitis amurensis*, confers abiotic stress tolerance in *Arabidopsis*.  
809 *Plant Cell Rep*. 2016 Mar;35(3):655-666.

810 [32] 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. *Plant Cell*. 2008 Nov;20(11):3122-3135.

[33] 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 May 1;17(9):1175-1187.

[34] Li X, Qian Q, Fu Z, Wang Y, Xiong G, Zeng D, Wang X, Liu X, Teng S, Hiroshi F, Yuan M, Luo D, Han B, Li J. Control of tillering in rice. *Nature*. 2003 Apr 10;422(6932):618-621.

[35] 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 Aug 7;10(8):e1004519.

[36] Fukazawa J, Teramura H, Murakoshi S, Nasuno K, Nishida N, Ito T, Yoshida M, Kamiya Y, Yamaguchi S, Takahashi Y. DELLAs function as coactivators of GAI-ASSOCIATED FACTOR1 in regulation of gibberellin homeostasis and signaling in Arabidopsis. *Plant Cell*. 2014 Jul;26(7):2920-2938.

[37] Yoshida H, Hirano K, Sato T, Mitsuda N, Nomoto M, Maeo K, Koketsu E, Mitani R, Kawamura M, Ishiguro S, Tada Y, Ohme-Takagi M, Matsuoka M, Ueguchi-Tanaka M. DELLA protein functions as a transcriptional activator through the DNA binding of the indeterminate domain family proteins. *Proc Natl Acad Sci U S A*. 2014 May 27;111(21):7861-7866.

[38] Liu W, Kohlen W, Lillo A, Op den Camp R, Ivanov S, Hartog M, Limpens E, Jamil M, Smaczniak C, Kaufmann K, Yang WC, Hooiveld GJ, Charnikhova T, Bouwmeester HJ, Bisseling T, Geurts R. Strigolactone biosynthesis in *Medicago truncatula* and rice requires the symbiotic GRAS-type transcription factors NSP1 and NSP2. *Plant Cell*. 2011 Oct;23(10):3853-3865.

[39] Sun L, Li X, Fu Y, Zhu Z, Tan L, Liu F, Sun X, Sun X, Sun C. GS6, a member of the GRAS gene family, negatively regulates grain size in rice. *J Integr Plant Biol*. 2013 Oct;55(10):938-949.

[40] Tong H, Jin Y, Liu W, Li F, Fang J, Yin Y, Qian Q, Zhu L, Chu C. DWARF AND LOW-TILLERING, a new member of the GRAS family, plays positive roles in brassinosteroid signaling in rice. *Plant J*. 2009 Jun;58(5):803-816.

[41] Itoh H, Sasaki A, Ueguchi-Tanaka M, Ishiyama K, Kobayashi M, Hasegawa Y, Minami E, Ashikari M, Matsuoka M. Dissection of the phosphorylation of rice DELLA protein, SLENDER RICE1. *Plant Cell Physiol*. 2005 Aug;46(8):1392-1399.

- [42] Li H , Payne W A , Michels G J , et al. Reducing plant abiotic and biotic stress: Drought and attacks of greenbugs, corn leaf aphids and virus disease in dryland sorghum. *Environmental & Experimental Botany*, 2008, 63(1-3):305-316.
- [43] Ikeda A, Ueguchi-Tanaka M, Sonoda Y, Kitano H, Koshioka M, Futsuhara Y, Matsuoka M, Yamaguchi J. slender rice, a constitutive gibberellin response mutant, is caused by a null mutation of the SLR1 gene, an ortholog of the height-regulating gene GAI/RGA/RHT/D8. *Plant Cell*. 2001 May;13(5):999-1010.
- [44] Stuurman J, Jäggi F, Kuhlemeier C. Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells. *Genes Dev*. 2002 Sep 1;16(17):2213-2218.
- [45] Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN. The SHORT-ROOT gene controls radial patterning of the Arabidopsis root through radial signaling. *Cell*. 2000 May 26;101(5):555-567.
- [46] Lu J, Wang T, Xu Z, Sun L, Zhang Q. Genome-wide analysis of the GRAS gene family in *Prunus mume*. *Mol Genet Genomics*. 2015 Feb;290(1):303-317.
- [47] Huang W, Xian Z, Kang X, Tang N, Li Z. Genome-wide identification, phylogeny and expression analysis of GRAS gene family in tomato. *BMC Plant Biol*. 2015 Aug 25;15:209.
- [48] 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 Mar 30;7:353.
- [49] 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 Jun 24;17(7):1004.
- [50] Fan S, Zhang D, Gao C, Zhao M, Wu H, Li Y, Shen Y, Han M. Identification, Classification, and Expression Analysis of GRAS Gene Family in *Malus domestica*. *Front Physiol*. 2017 Apr 28;8:253.
- [51] 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 Sep 28;12(9):e0185418.
- [52] Wang YX, Liu ZW, Wu ZJ, Li H, Wang WL, Cui X, Zhuang J. Genome-wide identification and expression analysis of GRAS family transcription factors in tea plant (*Camellia sinensis*). *Sci Rep*. 2018 Mar 2;8(1):3949.

871 [53] Zhang B, Liu J, Yang ZE, Chen EY, Zhang CJ, Zhang XY, Li FG. Genome-wide analysis of  
872 GRAS transcription factor gene family in *Gossypium hirsutum* L. BMC Genomics. 2018 May  
873 9;19(1):348.

874 [54] Liu B, Sun Y, Xue J, Jia X, Li R. Genome-wide characterization and expression analysis of  
875 GRAS gene family in pepper (*Capsicum annuum* L.). PeerJ. 2018 May 29;6:e4796.

876 [55] Zeng X, Ling H, Chen X, Guo S. Genome-wide identification, phylogeny and function analysis  
877 of GRAS gene family in *Dendrobium catenatum* (Orchidaceae). Gene. 2019 Jul 15;705:5-15.

878 [56] Quan S, Niu J, Zhou L, Xu H, Ma L, Qin Y. Genome-wide Identification, Classification,  
879 Expression and Duplication Analysis of GRAS Family Genes in *Juglans regia* L. Sci Rep. 2019 Aug  
880 12;9(1):11643.

881 [57] Liu M, Huang L, Ma Z, Sun W, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-wide identification,  
882 expression analysis and functional study of the GRAS gene family in Tartary buckwheat  
883 (*Fagopyrum tataricum*). BMC Plant Biol. 2019 Aug 6;19(1):342.

884 [58] Guo P, Wen J, Yang J, Ke Y, Wang M, Liu M, Ran F, Wu Y, Li P, Li J, Du H. Genome-wide  
885 survey and expression analyses of the GRAS gene family in *Brassica napus* reveals their roles in  
886 root development and stress response. Planta. 2019 Oct;250(4):1051-1072.

887 [59] Zhang H, Mi L, Xu L, Yu C, Li C, Chen C. Genome-wide identification, characterization,  
888 interaction network and expression profile of GRAS gene family in sweet orange (*Citrus sinensis*).  
889 Sci Rep. 2019 Feb 15;9(1):2156.

890 [60] To VT, Shi Q, Zhang Y, Shi J, Shen C, Zhang D, Cai W. Genome-Wide Analysis of the GRAS  
891 Gene Family in Barley (*Hordeum vulgare* L.). Genes (Basel). 2020 May 14;11(5):553.

892 [61] Shan Z, Luo X, Wu M, Wei L, Fan Z, Zhu Y. Genome-wide identification and expression of  
893 GRAS gene family members in cassava. BMC Plant Biol. 2020 Jan 29;20(1):46.

894 [62] Zhao Y, Li X, Chen W, et al. Whole-genome survey and characterization of MADS-box gene  
895 family in maize and sorghum. Plant Cell Tissue and Organ Culture, 2011, 105 (2),159-173.

896 [63] Kushwaha H, Gupta S, Singh VK, Rastogi S, Yadav D. Genome wide identification of Dof  
897 transcription factor gene family in sorghum and its comparative phylogenetic analysis with rice and  
898 Arabidopsis. Mol Biol Rep. 2011,38 (8),5037-5053.

899 [64] Zhang C, Bian M, Yu H, Liu Q, Yang Z. Identification of alkaline stress-responsive genes of  
900 CBL family in sweet sorghum (*Sorghum bicolor* L.). Plant Physiol Biochem. 2011,49 (11),1306-

901 1312.

902 [65] Yan HW, Hong L, Zhou YQ, Jiang HY, Zhu SW, Fan J, Cheng BJ. A genome-wide analysis of  
903 the ERF gene family in sorghum. *Genet Mol Res.* 2013,21,12 (2),2038-2055.

904 [66] Chang JZ, Yan FX, Qiao LY, Zheng J, Zhang FY, Liu QS. Genome-wide identification and  
905 expression analysis of SBP-box gene family in *Sorghum bicolor* L. *Yi Chuan.* 2016,20,38 (6),569-  
906 580.

907 [67] Agaraju M, Reddy PS, Kumar SA, Kumar A, Rajasheker G, Rao DM, Kavi Kishor PB.  
908 Genome-wide identification and transcriptional profiling of small heat shock protein gene family  
909 under diverse abiotic stress conditions in *Sorghum bicolor* (L.). *Int J Biol Macromol.*  
910 2020,1,142,822-834.

911 [68] Agaraju M, Kumar SA, Reddy PS, Kumar A, Rao DM, Kavi Kishor PB. Genome-scale  
912 identification, classification, and tissue specific expression analysis of late embryogenesis abundant  
913 (LEA) genes under abiotic stress conditions in *Sorghum bicolor* L. *PLoS One.* 2019,16,14  
914 (1),e0209980.

915 [69] Sanjari S, Shirzadian-Khorramabad R, Shobbar ZS, Shahbazi M. Systematic analysis of NAC  
916 transcription factors' gene family and identification of post-flowering drought stress responsive  
917 members in sorghum. *Plant Cell Rep.* 2019,38 (3),361-376.

918 [70] Lu X, Liu W, Xiang C, Li X, Wang Q, Wang T, Liu Z, Zhang J, Gao L, Zhang W. Genome-  
919 Wide Characterization of GRAS Family and Their Potential Roles in Cold Tolerance of Cucumber  
920 (*Cucumis sativus* L.). *Int J Mol Sci.* 2020 May 29;21(11):3857.

921 [71] Song XM, Liu TK, Duan WK, Ma QH, Ren J, Wang Z, Li Y, Hou XL. Genome-wide analysis  
922 of the GRAS gene family in Chinese cabbage (*Brassica rapa* ssp. *pekinensis*). *Genomics.* 2014  
923 Jan;103(1):135-146.

924 [72] Zhang H, Cao Y, Shang C, Li J, Wang J, Wu Z, Ma L, Qi T, Fu C, Bai Z, Hu B. Genome-wide  
925 characterization of GRAS family genes in *Medicago truncatula* reveals their evolutionary dynamics  
926 and functional diversification. *PLoS One.* 2017 Sep 25;12(9):e0185439.

927 [73] Tripathi P, Rabara RC, Langum TJ, Boken AK, Rushton DL, Boomsma DD, Rinerson CI,  
928 Rabara J, Reese RN, Chen X, Rohila JS, Rushton PJ. The WRKY transcription factor family in  
929 *Brachypodium distachyon*. *BMC Genomics.* 2012 Jun 22;13:270.

930 [74] "Shabalina S A, Ogurtsov A Y, Spiridonov A N, et al. Distinct patterns of expression and

931 evolution of intronless and intron-containing mammalian  
 932 genes . Molecular Biology and Evolution, 2010, 27( 8) : 1745-1749. "

933 [75] "Sang Y M, Liu Q, Lee J, et al. Expansion of amphibian intronless interferons revises the  
 934 paradigm for interferon evolution and functional diversity .  
 935 Scientific Reports, 2016, 6: 29072. "

936 [76] "Jain M, Khurana P, Tyagi A K, et al. Genome - wide analysis of intronless genes in  
 937 rice and Arabidopsis . Functional and Integrative  
 938 Genomics, 2008, 8( 1) : 69-78. "

939 [77] Chen F , Hu Y , Vannozzi A , Wu KC , Cai HY , Qin Y , Mullis A , Lin ZG , Zhang LS. The  
 940 WRKY Transcription Factor Family in Model Plants and Crops. Critical Reviews in Plant Sciences,  
 941 2018,36 (5),1-25.

942 [78] Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism.  
 943 Genetics. 1989 Nov;123(3):585-95. PMID: 2513255; PMCID: PMC1203831.

944 [79] Sudhir K , Glen S , Koichiro T . MEGA7: Molecular Evolutionary Genetics Analysis Version  
 945 7.0 for Bigger Datasets[J]. Molecular Biology & Evolution, 2016(7):1870.

946 [80] Swain MS, Reid J, Kamiya Y. Gibberellins are required for embryo and seed development in  
 947 pea. Plant J. 2002;12(6):1329–1338.

948 [81] Atchley WR, Terhalle W, Dress A. Positional dependence, cliques, and predictive motifs in the  
 949 bHLH protein domain. J Mol Evol. 1999 May;48(5):501-516.

950 [82] Li S, Zhao Y, Zhao Z, Wu X, Sun L, Liu Q, Wu Y. Crystal Structure of the GRAS Domain of  
 951 SCARECROW-LIKE7 in *Oryza sativa*. Plant Cell. 2016 May;28(5):1025-1034.

952 [83] Sun X, Rikkerink EH, Jones WT, Uversky VN. Multifarious roles of intrinsic disorder in  
 953 proteins illustrate its broad impact on plant biology. Plant Cell. 2013 Jan;25(1):38-55.

954 [84] Abarca D, Pizarro A, Hernández I, Sánchez C, Solana SP, Del Amo A, Carneros E, Díaz-Sala  
 955 C. The GRAS gene family in pine: transcript expression patterns associated with the maturation-  
 956 related decline of competence to form adventitious roots. BMC Plant Biol. 2014 Dec 30;14:354.

957 [85] Aubourg S, Kreis M, Lecharny A. The DEAD box RNA helicase family in *Arabidopsis thaliana*.  
 958 Nucleic Acids Res. 1999 Jan 15;27(2):628-636.

959 [86] Jain M, Tyagi AK, Khurana JP. Genome-wide analysis, evolutionary expansion, and expression  
 960 of early auxin-responsive SAUR gene family in rice (*Oryza sativa*). Genomics. 2006 Sep;88(3):360-

371.

[87] Liu X, Widmer A. Genome-wide comparative analysis of the GRAS gene family in *Populus*, *Arabidopsis* and *Rice*. *Plant Mol Biol Rep*. 2014;32(6):1129-1145.

[88] Zhang D, Iyer LM, Aravind L. Bacterial GRAS domain proteins throw new light on gibberellic acid response mechanisms. *Bioinformatics*. 2012 Oct 1;28(19):2407-2411.

[89] Freeling M. Bias in plant gene content following different sorts of duplication: tandem, whole-genome, segmental, or by transposition. *Annu Rev Plant Biol*. 2009;60:433-453.

[90] Gremski K, Ditta G, Yanofsky MF. The HECATE genes regulate female reproductive tract development in *Arabidopsis thaliana*. *Development*. 2007 Oct;134(20):3593-3601.

[91] Zhang J, Luo T, Wang W, Cao T, Li R, Lou Y. Silencing OsSLR1 enhances the resistance of rice to the brown planthopper *Nilaparvata lugens*. *Plant Cell Environ*. 2017 Oct;40(10):2147-2159.

[92] Lin Z, Yan J, Su J, Liu H, Hu C, Li G, Wang F, Lin Y. Novel OsGRAS19 mutant, D26, positively regulates grain shape in rice (*Oryza sativa*). *Funct Plant Biol*. 2019 Aug;46(9):857-868.

[93] Kaló P, Gleason C, Edwards A, Marsh J, Mitra RM, Hirsch S, Jakab J, Sims S, Long SR, Rogers J, Kiss GB, Downie JA, Oldroyd GE. Nodulation signaling in legumes requires NSP2, a member of the GRAS family of transcriptional regulators. *Science*. 2005 Jun 17;308(5729):1786-1789.

[94] Wang Y, Deng D. Molecular basis and evolutionary pattern of GA-GID1-DELLA regulatory module. *Mol Genet Genomics*. 2014 Feb;289(1):1-9.

[95] Hu Y, Zhou L, Huang M, He X, Yang Y, Liu X, Li Y, Hou X. Gibberellins play an essential role in late embryogenesis of *Arabidopsis*. *Nat Plants*. 2018 May;4(5):289-298.

[96] Martí C, Orzáez D, Ellul P, Moreno V, Carbonell J, Granell A. Silencing of DELLA induces facultative parthenocarpy in tomato fruits. *Plant J*. 2007 Dec;52(5):865-876.

[97] Hu J, Israeli A, Ori N, Sun TP. The Interaction between DELLA and ARF/IAA Mediates Crosstalk between Gibberellin and Auxin Signaling to Control Fruit Initiation in Tomato. *Plant Cell*. 2018 Aug;30(8):1710-1728.

[98] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997, 1, 25 (17), 3389-3402.

[99] Liu M, Ma Z, Wang A, Zheng T, Huang L, Sun W, Zhang Y, Jin W, Zhan J, Cai Y, Tang Y, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-Wide Investigation of the Auxin Response Factor Gene

991 Family in Tartary Buckwheat (*Fagopyrum tataricum*). *Int J Mol Sci.* 2018, 9, 19 (11), 3526.  
 992 [100] Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity  
 993 searching. *Nucleic Acids Res.* 2011, 39 (Web Server issue), W29-W37.  
 994 [101] Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer EL. The Pfam protein  
 995 families database. *Nucleic Acids Res.* 2000, 1, 28 (1), 263-266.  
 996 [102] Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nucleic Acids*  
 997 *Res.* 2018, 4, 46 (D1), D493-D496.  
 998 [103] Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and  
 999 ClustalX. *Curr Protoc Bioinformatics.* 2002, Chapter 2: Unit 2. 3.  
 1000 [104] Guo AY, Zhu QH, Chen X, Luo JC. [GSDS: a gene structure display server]. *Yi Chuan.* 2007,  
 1001 29 (8), 1023-1026.  
 1002 [105] Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS.  
 1003 MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 2009, 37 (Web Server  
 1004 issue), W202-W208.  
 1005 [106] Xie T, Chen C, Li C, Liu J, Liu C, He Y. Genome-wide investigation of WRKY gene family  
 1006 in pineapple: evolution and expression profiles during development and stress. *BMC Genomics.*  
 1007 2018, 25, 19 (1), 490.  
 1008 [107] Liu M, Ma Z, Sun W, Huang L, Wu Q, Tang Z, Bu T, Li C, Chen H. Genome-wide analysis  
 1009 of the NAC transcription factor family in Tartary buckwheat (*Fagopyrum tataricum*). *BMC*  
 1010 *Genomics.* 2019, 6, 20 (1), 113.  
 1011 [108] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA.  
 1012 Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009, 19 (9), 1639-1645.  
 1013 [109] Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger  
 1014 JC, Paterson AH. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and  
 1015 collinearity. *Nucleic Acids Res.* 2012, 40 (7), e49.  
 1016 [110] Sudhakar Reddy P, Srinivas Reddy D, Sivasakthi K, Bhatnagar-Mathur P, Vadez V, Sharma  
 1017 KK. Evaluation of Sorghum [*Sorghum bicolor* (L.)] Reference Genes in Various Tissues and under  
 1018 Abiotic Stress Conditions for Quantitative Real-Time PCR Data Normalization. *Front Plant Sci.*  
 1019 2016, 25, 7, 529.  
 1020 [111] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time

quantitative PCR and the 2<sup>-</sup>(-Delta Delta C (T)) Method. Methods. 2001, 25 (4), 402-408.

**Figure legends:**

**Figure 1** Unrooted phylogenetic tree representing relationships among GRAS domains of *S. bicolor*, *Arabidopsis* and rice. The phylogenetic trees were derived using the NJ method in MEGA7.0. The tree shows the 13 phylogenetic subfamilies marked with red font on a white background. GRAS proteins from *Arabidopsis* and *Oryza sativa* have the prefix ‘At’ and ‘Os’, respectively.

**Figure 2** Multiple sequence alignment of the GRAS domains of the members of 13 phylogenetic subfamilies of the *S. bicolor* GRAS protein family. The scheme at the top depicts the locations and boundaries of the LHR I, VHIID, LHR II, PFYRE, and SAW regions within the GRAS domain.

**Figure 3** Phylogenetic relationship, gene-structure analysis, and motif distributions of *S. bicolor* GRAS genes.

**A** Phylogenetic tree was constructed by the NJ method with 1000 replicates on each node.

**B** Exons and introns are indicated by green rectangles and gray lines, respectively.

**C** Amino acid motifs in the *S. bicolor* GRAS proteins (1–10) are represented by colored boxes.

The black lines indicate relative protein lengths.

**Figure 4** Schematic representations of the chromosomal distribution of the *S. bicolor* GRAS genes. Vertical bars represent the chromosomes of *S. bicolor*. The chromosome number is indicated to the left of each chromosome. The scale on the left represents chromosome length.

**Figure 5** Schematic representations of the chromosomal distribution and segmental duplication relationships of *S. bicolor* GRAS genes. Colored lines indicate all synteny blocks in the *S. bicolor* genome and the red lines indicate duplicated GRAS gene pairs. The chromosome number is indicated at the bottom of each chromosome.

**Figure 6** Synteny analyses of the GRAS genes between *S. bicolor* and six representative plant species. Gray lines on the background indicate the collinear blocks within *S. bicolor* and other plant genomes; red lines highlight the syntenic *S. bicolor* GRAS gene pairs.

**Figure 7** Phylogenetic relationships and motif compositions of the *S. bicolor* GRAS proteins with six different plant species (*Arabidopsis thaliana*, *Capsicum annuum*, *Solanum lycopersicum*, *Brachypodium distachyon*, *Oryza sativa* subsp. *indica*, *Zea mays*). Outer panel: Unrooted phylogenetic tree constructed using Geneious R11 with the NJ method. Innermost panel:

Distribution of the conserved motifs in GRAS proteins. The differently colored boxes represent different motifs and their positions in each GRAS protein sequence. The sequence information for each motif is provided in [Additional file 3: Table S3](#).

**Figure 8** Tissue-specific expression of 13 *S. bicolor* GRAS genes and the correlation between their expression patterns.

**A** Expression patterns of 13 *S. bicolor* GRAS genes in anther, style, leaf, root, stem and fruit organs were examined by qRT-PCR. Error bars were obtained from three measurements. Lowercase letters above the bars indicate significant differences ( $\alpha = 0.05$ , LSD) among the treatments.

**B** Positive number: positively correlated; negative number: negatively correlated. Red numbers indicate a significant correlation at the 0.05 level.

**Figure 9** Expression of 13 *S. bicolor* GRAS genes under different abiotic stresses (strong UV radiation, flooding, PEG, NaCl, heat and cold treatments) at the seedling stage. Error bars were obtained from three measurements. Lowercase letters above the bars indicate significant differences ( $\alpha = 0.05$ , LSD) among the treatments.

**A** Expression patterns of 13 *S. bicolor* GRAS genes in leaf, root and stem in were examined by qRT-PCR.

**B** Positive number: positively correlated; negative number: negatively correlated. The red numbers indicate a significant correlation at the 0.05 level.

**C** Gibberellin (GA) content in the different tissues.

#### **Additional files:**

**Additional file 1: Table S1.** List of the 81 *S. bicolor* GRAS genes identified in this study.

**Additional file 2: Table S2.** Subfamilies and protein sequences of *Arabidopsis* and rice.

**Additional file 3: Table S3.** Analysis and distribution of the conserved motifs of GRAS proteins.

**Additional file 4: Table S4.** Tandem duplication events of *S. bicolor* GRAS genes.

**Additional file 5: Table S5.** The 14 pairs of segmental duplicates in *S. bicolor* GRAS genes.

**Additional file 6: Table S6.** One-to-one orthologous relationships between *S. bicolor* and other plants.

**Additional file 7: Table S7.** Results of Tajima's D neutrality test.

**Additional file 8: Table S8.** Primer sequences for qRT-PCR.

1081

1082    **Abbreviations:**

1083    *At: Arabidopsis thaliana*

1084    *Os: Oryza sativa*

1085    *AtGRAS: Arabidopsis thaliana GRAS*

1086    *OsGRAS: Oryza sativa GRAS*

1087    *SbGRAS: Sorghum bicolor GRAS*

1088    GAPDH: Glyceraldehyde-3-phosphate dehydrogenase

1089    qRT-PCR: Quantitative real-time polymerase chain reaction

1090    TF: Transcription factor

1091    CDS: Coding sequence

1092    HMM: Hidden Markov Model

1093    pI: Isoelectric point

1094    BLAST: Basic local alignment search tool

1095    LHR I: Leucine-heptad repeat I

1096    LHR II: Leucine-heptad repeat II

1097    SCR: SCARECROW

1098

1099

1100

1101

1102

1103

1104

1105

1106

1107

1108

1109

Figure 1

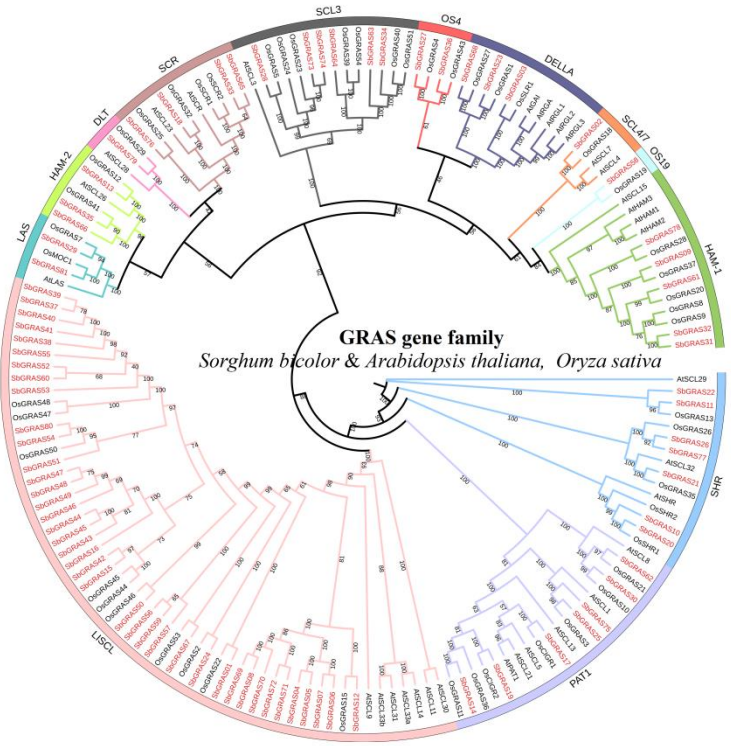
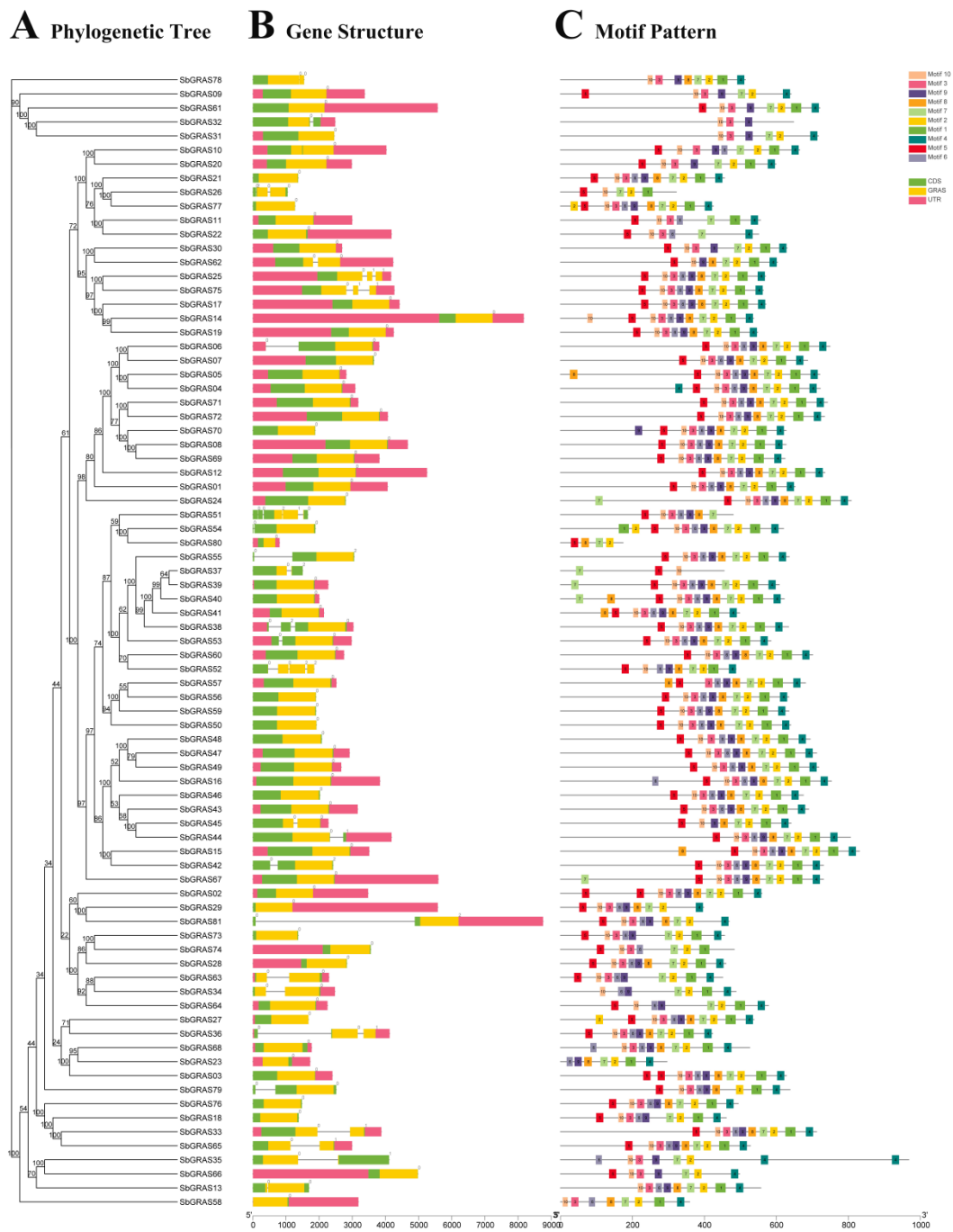


Figure 2



Figure 3



**Figure 4**

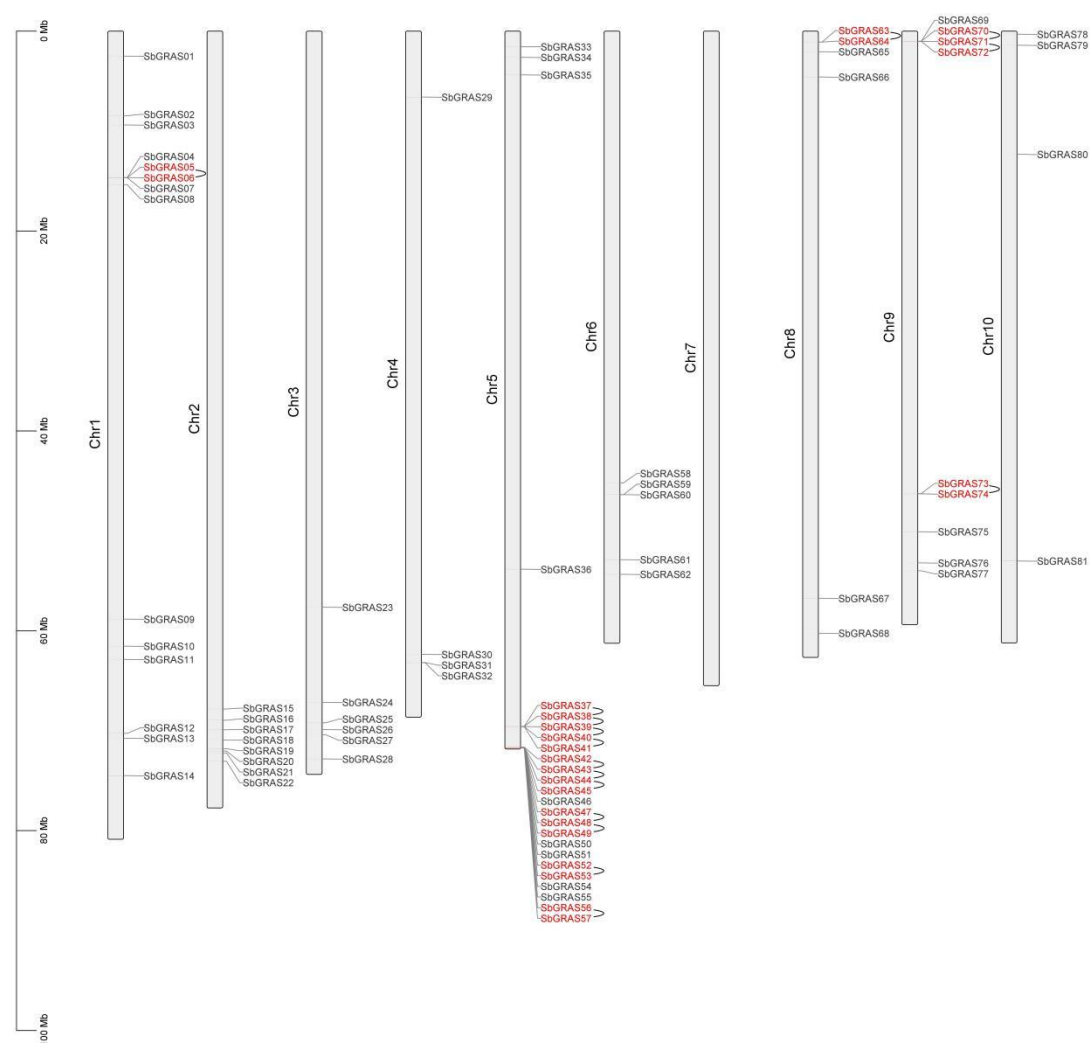
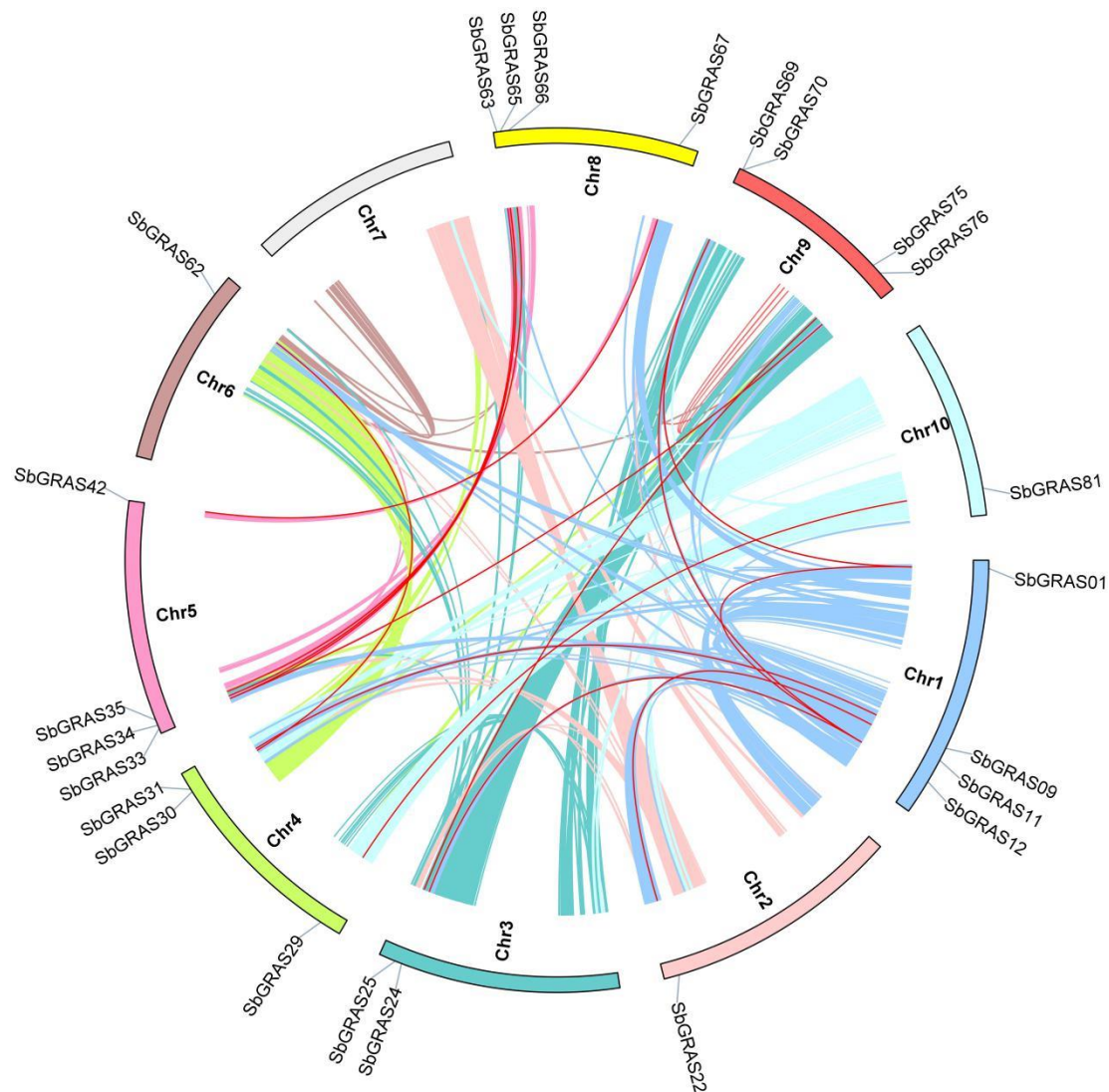


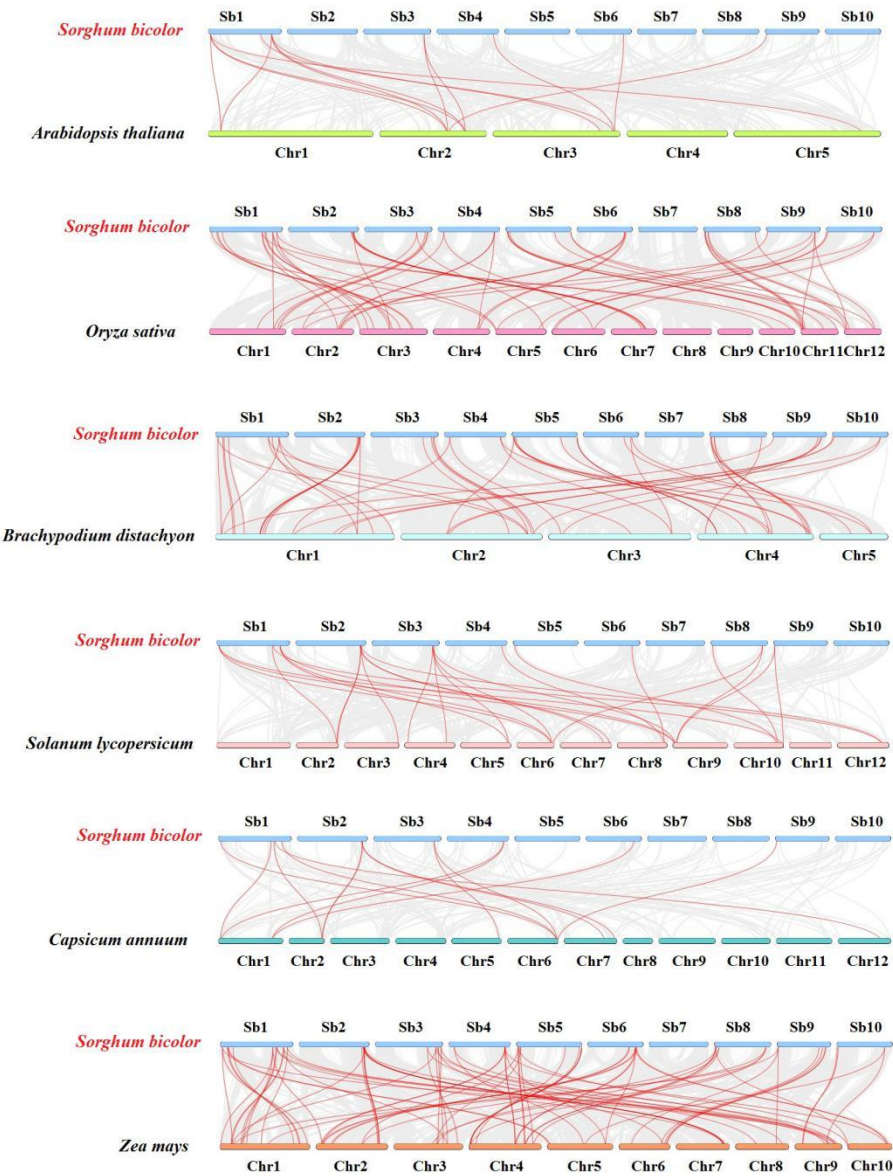
Figure 5



1173

1174

1175 **Figure 6**



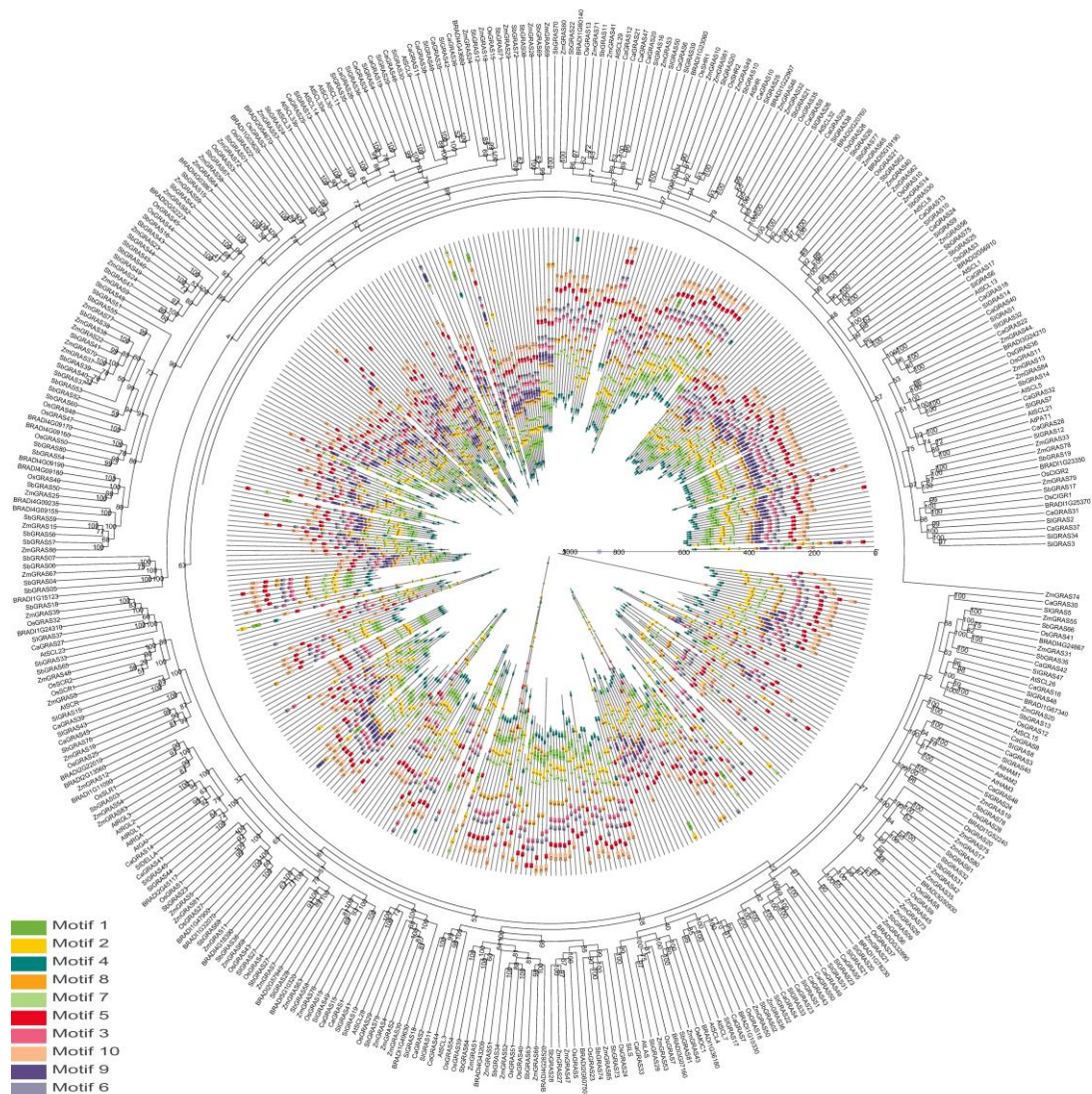
1176

1177

1178

1179

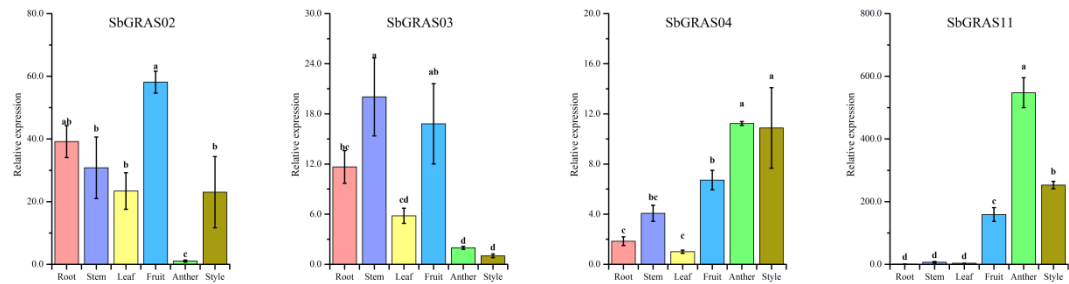
Figure 7



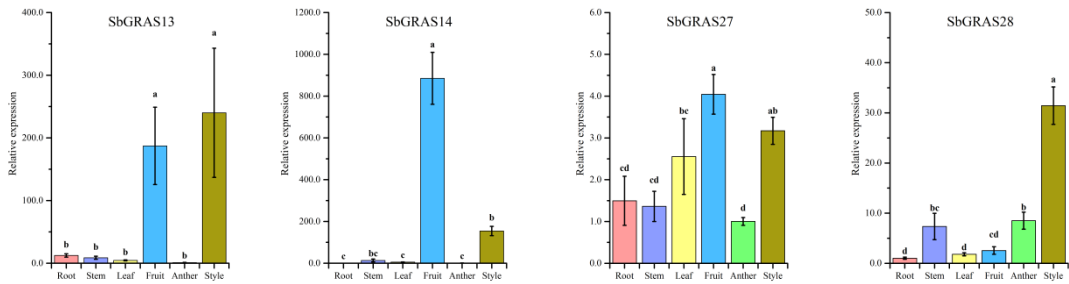
1193

1194

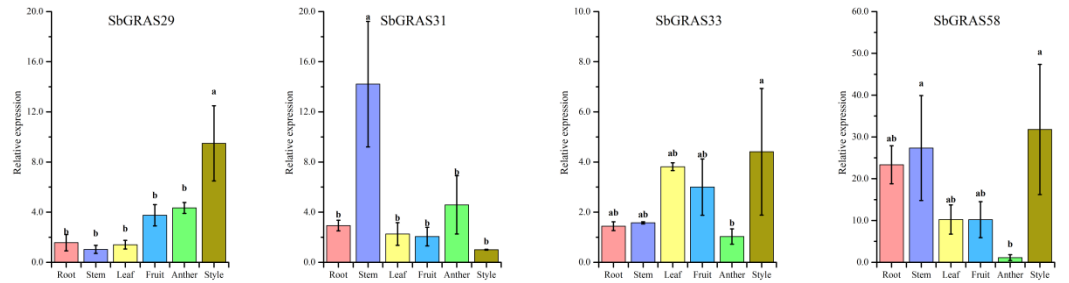
1195 **Figure 8A**



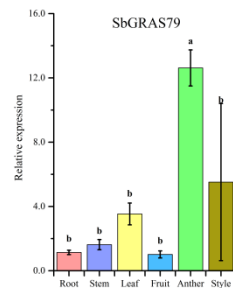
1196



1197



1198



1199

1200

1201

Figure 8B

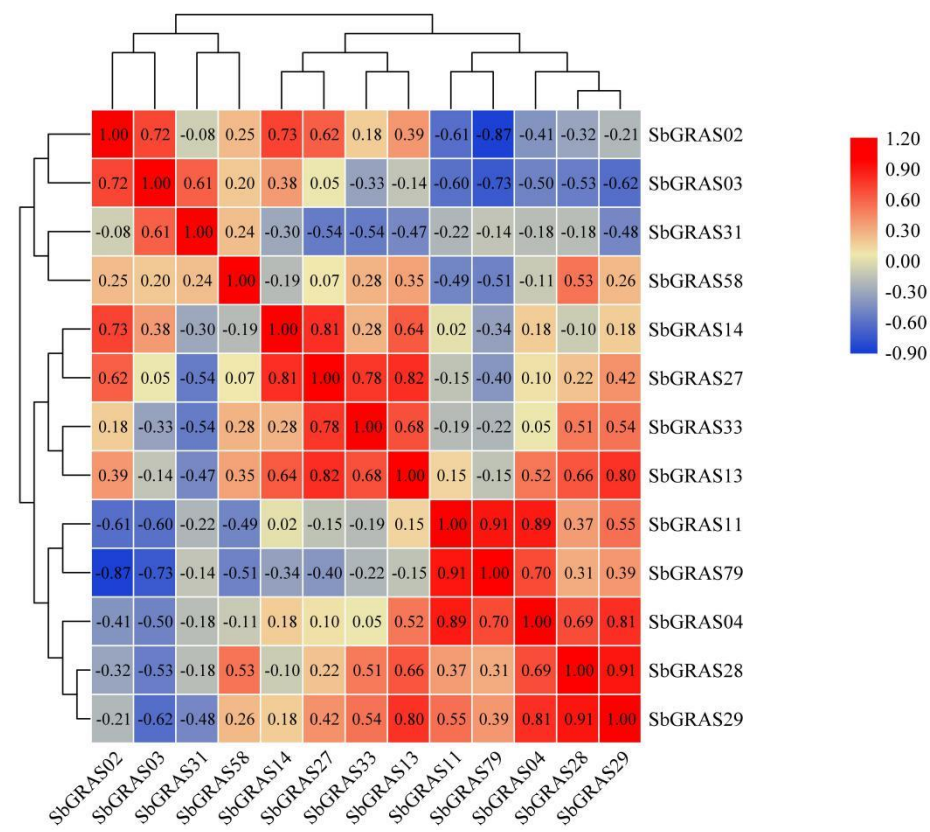
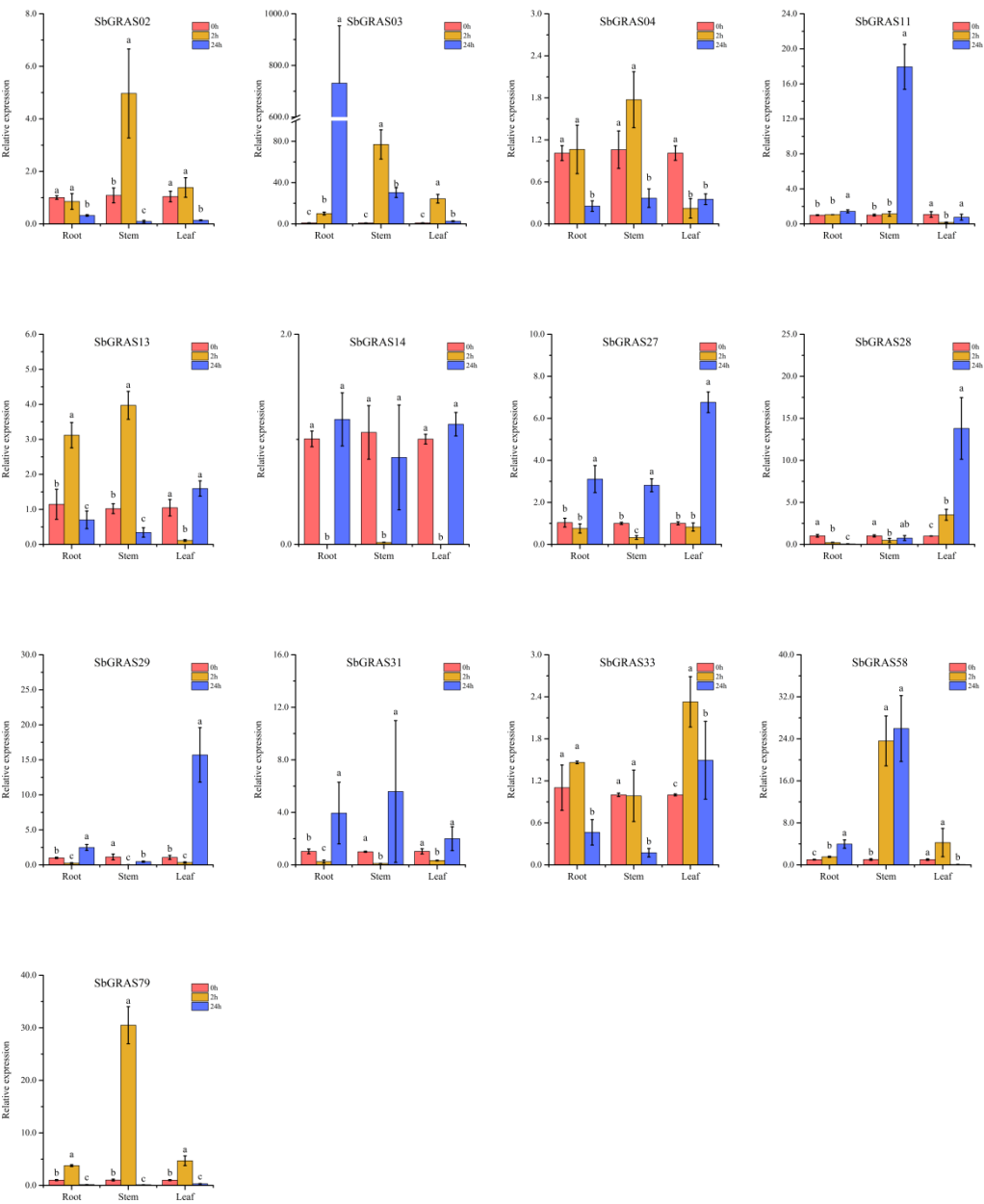


Figure 9A

UV



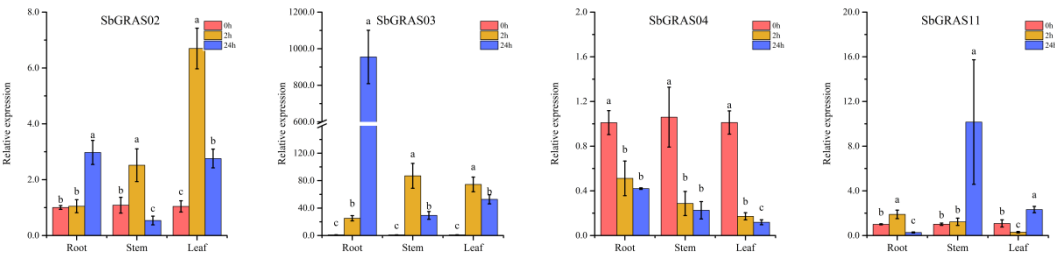
1229

1230

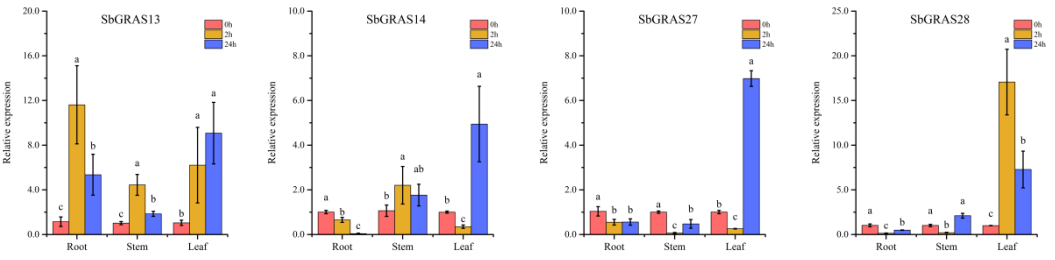
1231

# Flooding

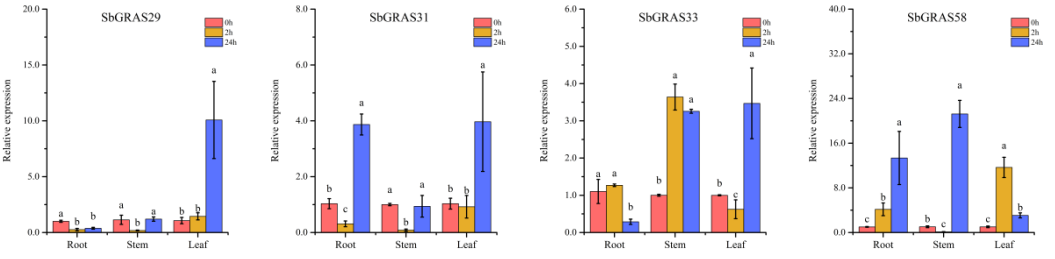
1232



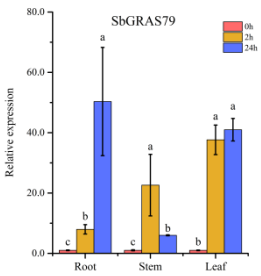
1233



1234



1235



1236

1237

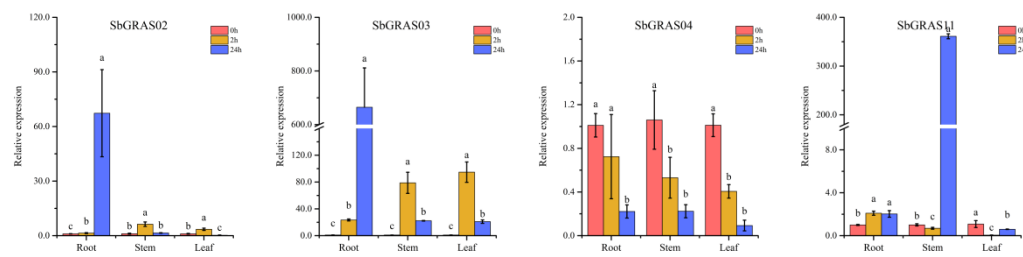
1238

1239

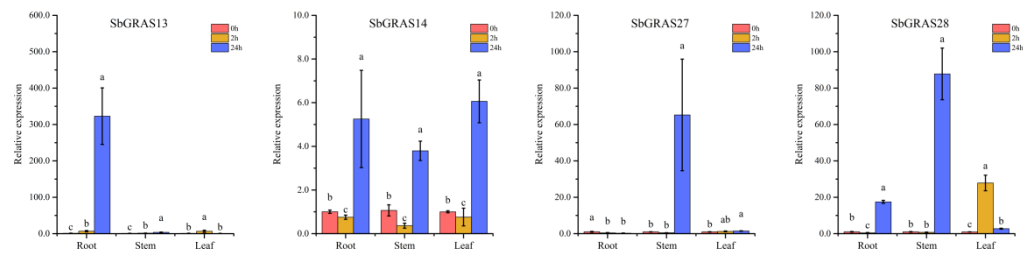
1240

PEG

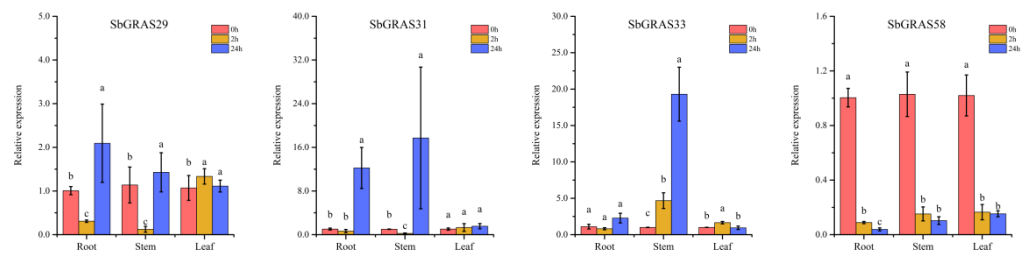
1241



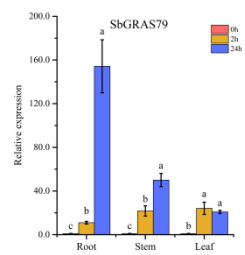
1242



1243



1244



1245

1246

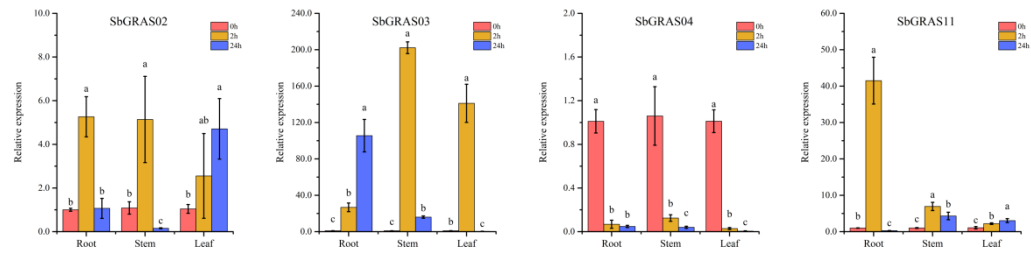
1247

1248

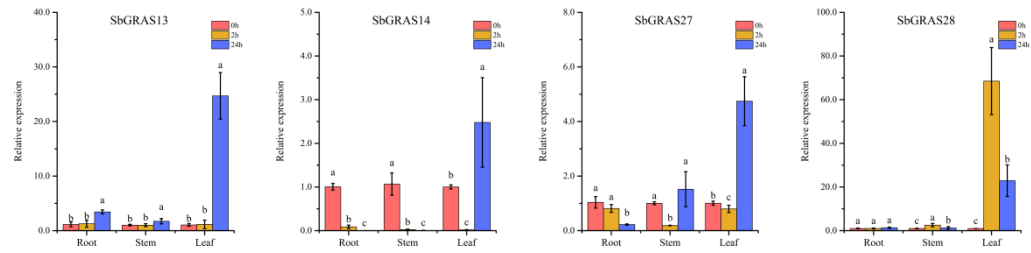
1249

NaCl

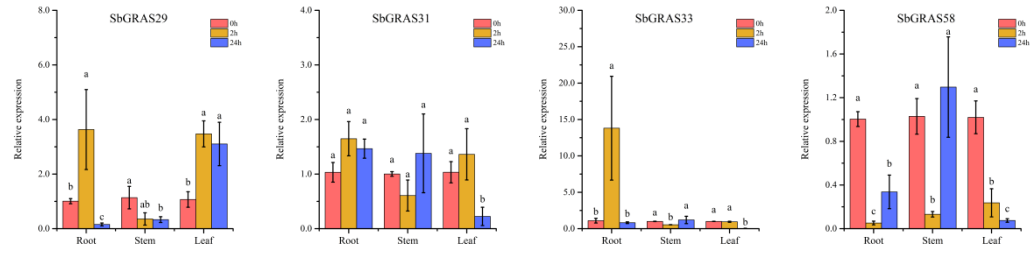
1250



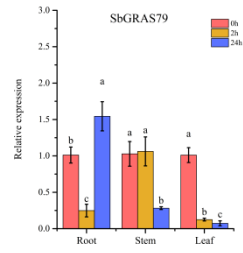
1251



1252



1253



1254

1255

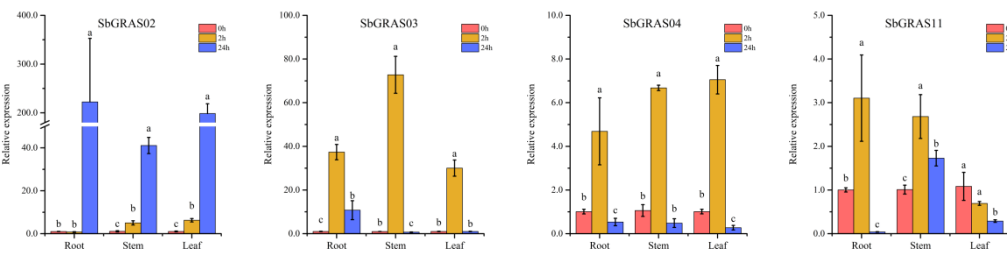
1256

1257

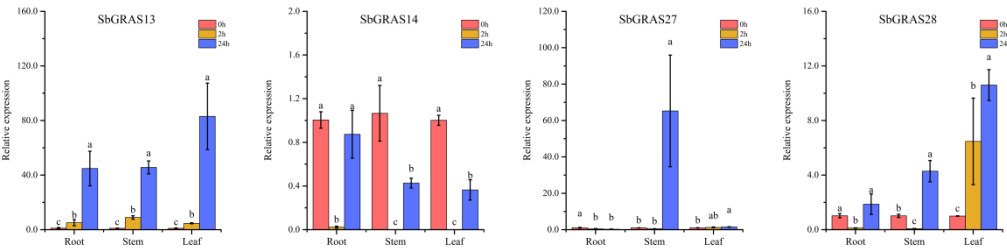
1258

Heat

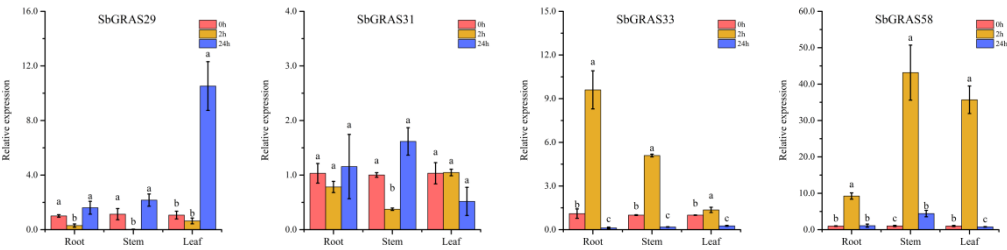
1259



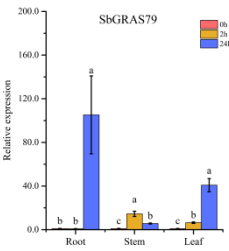
1260



1261



1262



1263

1264

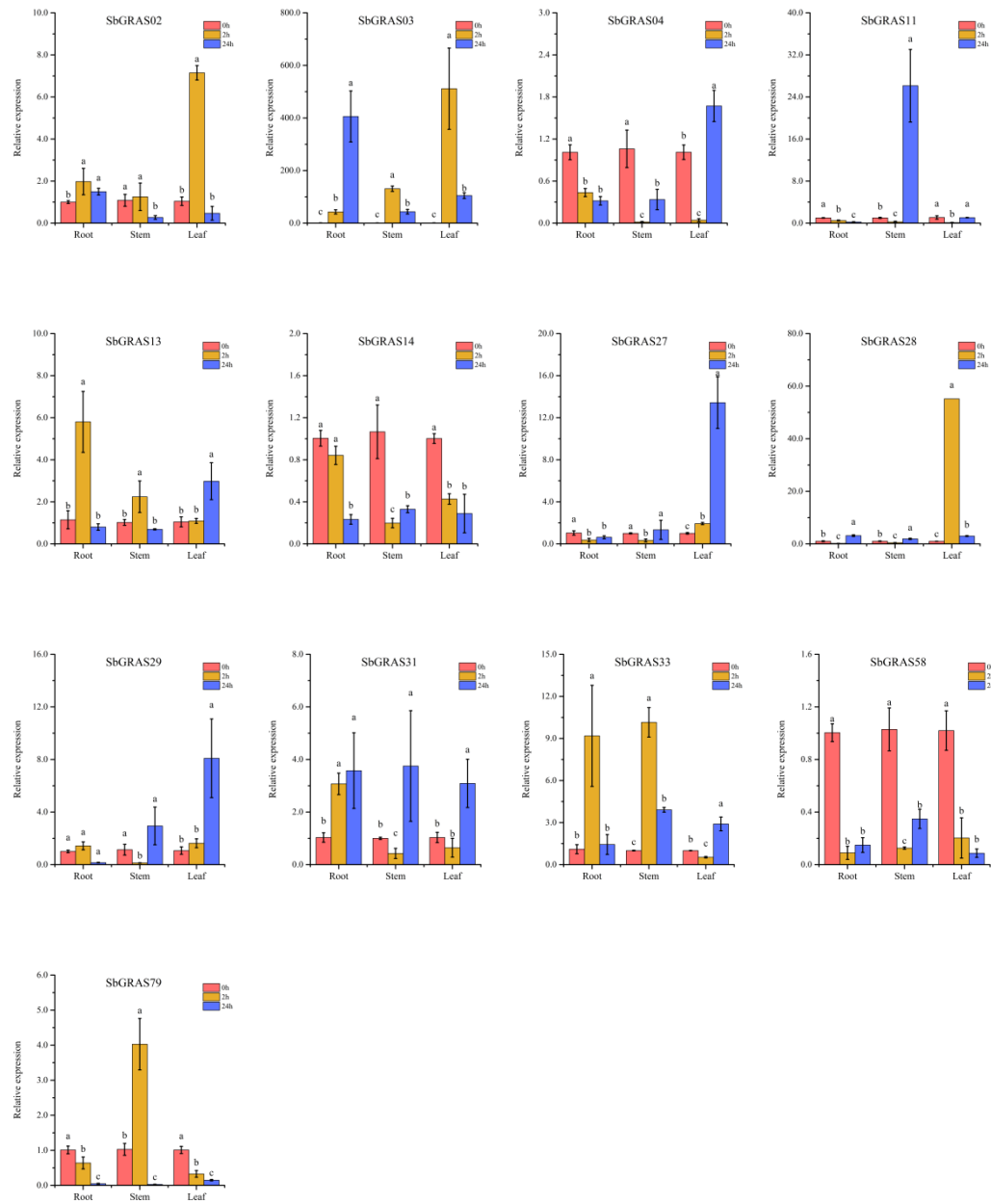
1265

1266

1267

Cold

1268



1271

1272

1273

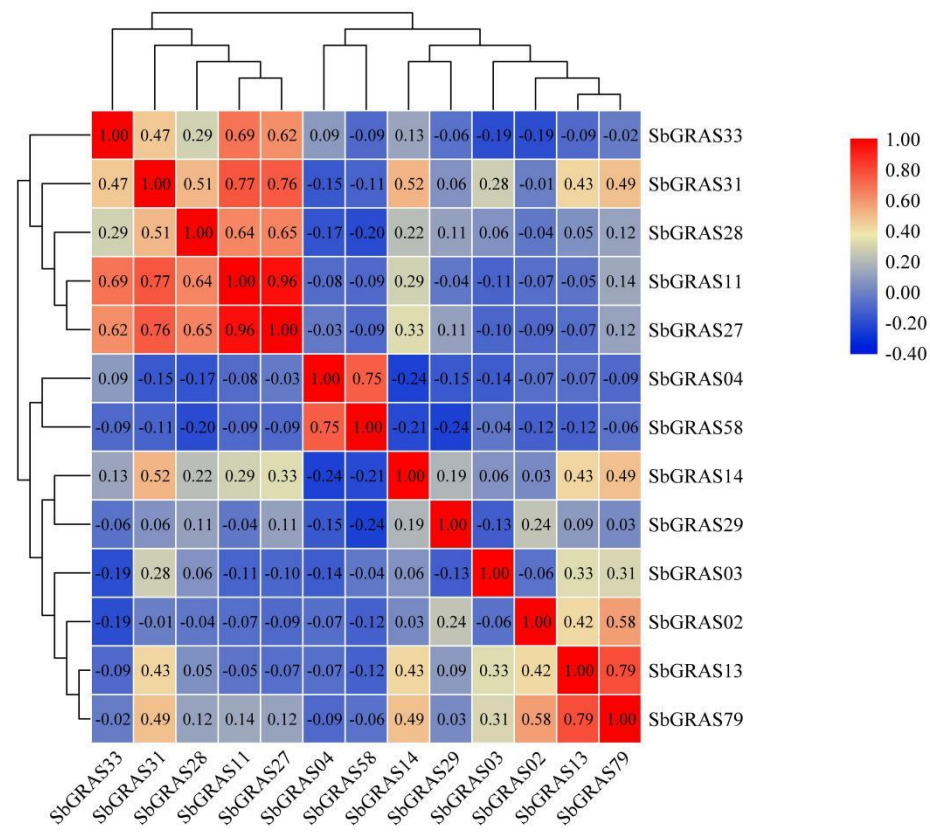


Figure 9C

