INTERNATIONAL JOURNAL OF AGRICULTURE & BIOLOGY
ISSN Print: 1560–8530; ISSN Online: 1814–9596
20–1133/2021/25–1–181–190
DOI: 10.17957/IJAB/15.1654
http://www.fspublishers.org

Full Length Article

Genome-Wide Dissection, Characterization, and Expression Profiling of Cotton GASA Genes Reveal their Importance in Regulating Abiotic Stresses

Muhammad Shaban[1], Aamir Hamid Khan[2], Etrat Noor[1], Abdul Qayyum[1], Waqas Malik[1], Muhammad Shehzad[3], Umar Akram[4], Ayesha Razzaq[1] and Muhammad Adnan Tabassum[5]

[1]*Genomics Lab, Department of Plant Breeding and Genetics, Bahauddin Zakariya University, Multan, Pakistan*
[2]*National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei 430070, P.R. China*
[3]*State Key Laboratory of Cotton Biology, Institute of Cotton Research, Chinese Academy of Agricultural Sciences, Anyang 455000, Henan, P.R. China*
[4]*Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Beijing, P.R. China*
[5]*Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of Education of China, Institutes of Agricultural Science and Technology Development, Yangzhou University, Jiangsu, China*
[*]For correspondence: raoqayyim@bzu.edu.pk, drshabaan@outlook.com
Received 08 August 2020; Accepted 21 September 2020; Published 10 December 2020

Abstract

Short amino acids constituting proteins of gibberellic acid stimulated Arabidopsis (GASA) gene family are widely implicated in plant growth, development and have potential in mitigating environmental stresses. There is limited information about functions of these genes in cotton. In the present study, a total of 116 GASA genes were screened from four species of cotton. During phylogenetic clustering, these genes were distributed into three groups based on their homology. Duplication analysis among three species of cotton revealed that segmental duplication events might be the possible reason for expansion and domestication of cultivated tetraploid cotton. Further, chromosomal distributions of GASA genes on cotton chromosomes were found uneven. The genes structure and motifs division pattern of *GhGASA* genes within same group was relatively conserved. Promoter regions analysis of *GhGASA* genes comprehend their involvement in a variety of plant mechanisms related to growth and survival against environmental stresses. In tissue-specific expression analysis, significantly higher induction of *GhGASA* genes in various tissues of upland cotton revealed their importance in development of these tissues. Additionally, differential expressions of *GhGASA* genes to multiple abiotic stresses, especially against salt and cold stresses predicted their potential roles in regulating these environmental cues. In conclusion, this is the first comprehensive study regarding identification and investigation of cotton GASA gene family. Data presented here provide important information for future elucidating and characterizing potential target GASA genes related to abiotic stress resistance in cotton.

Keywords: GASA genes; Cotton; Genome-wide; Tissue expression; Stress responses

Introduction

Cysteine-rich peptide (CRP) is a large group of proteins including thionines, lipid transfer proteins, defensins and GASA/Snakin. Each CRP class can be distinguished from other based on the orientation and number of cysteine residues in the primary sequence (Oliveira-Lima *et al*. 2017). Recently, CRP proteins have been reported extensively for their functions in plant development and manipulation of environmental stresses (Balaji and Smart 2012; Haruta *et al*. 2014; Ahmad *et al*. 2019; Li *et al*. 2019; Ahmad *et al*. 2020). CRP proteins constitute a large gene family and widely distributed in plants. The GASA (Gibberellic-acid stimulated Arabidopsis) a subfamily of CRP, is short amino acids, low molecular weights, mostly gibberellin regulated and is widely distributed in plants. It is comprised of three distinct domains, 18–29 residues comprising N-terminal domain, a C-terminal domain (called GASA domain) consisting of 12 conserved cysteine residues and an intermediate highly divergent region between C and N terminals (Herzog *et al*. 1995). The fact that number and position of C-terminal conserved residues remained same throughout the evolution in different species might suggests their key role in defining functions of this gene family (Ben-Nissan *et al*. 2004). Previous studies have reported that C-terminal GASA domain is essential for

To cite this paper: Shaban M, AH Khan, E Noor, A Qayyum, W Malik, M Shehzad, U Akram, A Razzaq, MA Tabassum (2021). Genome-wide dissection, characterization, and expression profiling of cotton GASA genes reveal their importance in regulating abiotic stresses. *Intl J Agric Biol* 25:181–190

determining antioxidant activity (Wigoda *et al*. 2006; Nahirnak *et al*. 2012b) and formation of disulfide bond during protein folding (Porto and Franco 2013).

GASA gene family plays important roles during different processes in plant life from seed germination to maturity. Studies based on expression pattern analysis suggest that GASA genes have specific spatial and temporal expression pattern and most of these are expressed in young tissues and actively growing organs (Peng *et al*. 2010; Nahirnak *et al*. 2016). These are not only the target genes responsible for specific functions but also act as regulatory proteins to monitor plants signaling for growth and stress responses (Ceserani *et al*. 2009; Zhang *et al*. 2009; Sun *et al*. 2013). Further, these genes control plant hormonal level and hormonal signaling network to fine tune different physiological processes in plants (Wang *et al*. 2009; Rubinovich *et al*. 2014). Additionally, transgenic studies, homology and expression analysis indicate substantial involvement of GASA genes in plant developmental processes by affecting various cellular processes (Shi *et al*. 1992; Aubert *et al*. 1998; Kotilainen *et al*. 1999; Fuente *et al*. 2006; Furukawa *et al*. 2006; Nahirnak *et al*. 2012a). More importantly, GASA family genes have preferential roles in floral development and regulation of floral timing (Muhammad *et al*. 2019). Further, GASA genes also modulate cell elongation (Ben-Nissan and Weiss 1996), root growth (Zimmermann *et al*. 2010) and stem development (Zhang *et al*. 2009) and fruits ripening (Moyano-Canete *et al*. 2013) in plants. Interestingly, some GASA family members have opposite functions related to flowering in plants, such as *AtGASA4* stimulates flowering (Roxrud *et al*. 2007) while *AtGASA5* inhibits flowering in Arabidopsis (Zhang *et al*. 2009).

In addition to plant growth and developmental functions of GASA genes, some members of this family are also involved in regulating plants stress responses. For example, increased anti-bacterial and anti-fungal activities were correlated with potato Snakin-1 and Snakin-2 proteins during *in vivo* experiments (Segura *et al*. 1999; Berrocal-Lobo *et al*. 2002; Almasia *et al*. 2008; Kovalskaya and Hammond 2009; Balaji and Smart 2012). Similarly, *CaSn* protein enhanced pepper resistance against root-knot nematode (Mao *et al*. 2011). In *Arabidopsis*, *AtGASA4* and *AtGASA5* substantially enhanced tolerance to heat stress by affecting BiP gene expression and regulating SA signaling pathway, respectively (Ko *et al*. 2007; Zhang and Wang 2011). Further, constitutive expression of *AtGASA14* increased tolerance to salinity by restricting accumulation of reactive oxygen species (ROS) (Sun *et al*. 2013). Likewise, overexpression of *FsGASA4* improves plant tolerance to abiotic stresses by enhancing SA level and induced expressions of SA signaling pathway genes (Alonso-Ramirez *et al*. 2009).

Gibberellic Acid-Stimulated Transcript 1 (GAST1) was the first GASA gene isolated and characterized in tomato during investigation of GA-deficient *gib* mutant (Shi *et al*. 1992). Later, with the advancement of sequencing technologies, more members of GASA genes were reported in diverse plant species including *Arabidopsis thaliana* (Herzog *et al*. 1995) wheat (Zhang *et al*. 2017), rice (Furukawa *et al*. 2006), maize (Zimmermann *et al*. 2010), apple (Fan *et al*. 2017), petunia (Ben-Nissan and Weiss 1996), potato (Nahirnak *et al*. 2016), soybean (Ahmad *et al*. 2019) and grapevine (Ahmad *et al*. 2020). GASA family constitutes a large number of genes in some species, for example, in soybean 37 GASA/Snakin genes were identified (Ahmad *et al*. 2019), similarly 16 in potato (Nahirnak *et al*. 2016), 15 in Arabidopsis (Fan *et al*. 2017), 14 in grapevine have been reported (Ahmad *et al*. 2020). The diversity and number of GASA/Snakin genes identified in remotely related species depicts their significance and suggest their essential roles in life of plants.

Cotton (*G. hirsutum*) is a naturally fiber and oil producing crop of huge importance for textile and oil industry of the world. The economy of many developing countries depends on the sustainable production of cotton. As cotton is a mesophytic plant, its growth and yield are severally affected by both biotic and abiotic stresses (Dabbert and Gore 2014). Based on the significance of GASA genes in regulating growth, development and responses to different environmental stresses in multiple plant species, GASA family was selected for systematic and comprehensive analysis in cotton. In this study, a total 116 putative GASA genes were screened from four species of cotton. Their phylogeny, synteny, motifs, structural features, *cis*-elements, spatiotemporal expressions in various tissues and under abiotic stresses (cold, heat, salinity) was investigated in details. Our results laid a foundation for further characterization of GASA family related to growth, development and responses to different environmental stresses.

Materials and Methods

Identification, sequence analysis and properties of GASA gene family in *Gossypium* spp.

All the reported Arabidopsis GASA family genes were downloaded (Fan *et al*. 2017). Afterwards, BLASTP programme (Zhu *et al*. 2017) with e-value $1e^{-10}$ was run online to find candidate GASA genes from four species of cotton including *G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii* using 15 Arabidopsis GASA genes as query. Subsequently, manual and online databases including NCBI conserved domain (https://ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) (Marchler-Bauer *et al*. 2015) Pfam (http://pfam.xfam.org/) (El-Gebali *et al*. 2019), and SMART (http://smart.embl-heidelberg.de/) (Letunic *et al*. 2015) were accessed to confirm GASA domain of Pfam (PF02704) from all the members of GASA family. Further, physio-chemical properties of GASA genes including theoretical molecular

weight, protein length and isoelectric point were determined through ExPASy web tool (http://web.expasy.org/) (Bjellqvist *et al*. 1994).

Chromosomal mapping, synteny and duplication analysis

Chromosomal positions of GASA genes were plotted on cotton chromosomes using CIRCOS software. Based on the gene position, the distribution of each GASA gene was analyzed. Further, homologous genes among *G. hirsutum*, *G. arboreum* and *G. raimondii* were screened using BLASTP program with similarity > 80% and alignment percentage > 80% compared to total length of proteins (Yang *et al*. 2008). The sequences of orthologous genes were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clusalo/). Subsequently, aligned sequences were submitted to PAL2NAL software (http://www.bork.embl.de/pal2nal/) (Suyama *et al*. 2006) for determination of synonymous (Ks), non-synonymous (Ka) substitution rates and their ratios among duplicated gene pairs. Separately, *GhGASA* genes were also mapped on chromosomes using Mapchart software (Version 2.0) (Voorrips 2002).

Phylogenetic tree construction

The phylogenetic tree between *GhGASA*, *GbGASA*, *GaGASA* and *GrGASA* and *Arabidopsis* GASA genes was constructed using their protein sequences. Firstly, multiple sequence alignment of protein sequences was generated using ClustalW software (Thompson *et al*. 2002). Later, this alignment file was submitted in MEGA 6.0 software (Tamura *et al*. 2013) for generation of phylogenetic tree. Neighbor-joining (NJ) method was adopted with parameters as follows, bootstrap values: 1000 replicate, pairwise deletion, Poisson correction and uniform rates.

Gene structure, motifs prediction and *cis*-elements analysis

Coding DNA sequence and corresponding Genomic sequence of each *GhGASA* gene was applied in online tool Gene Structure Display Server (GSDS, V.2) (http://gsds.cbi.pku.edu.cn/) (Hu *et al*. 2015) to construct gene model showing organization and number of exon-introns. The conserved motifs in *GhGASA* genes were predicted using MEME suits (Multiple Expectation Maximization for Motif Elicitation) (Bailey *et al*. 2015). Further, annotations of these motifs were obtained from Pfam database (http://pfam.xfam.org/) (El-Gebali *et al*. 2019). 1.5 kb upstream 5' flanking region of each *GhGASA* gene was retrieved from CottonFGD website (https://cottonfgd.org/analyze/) (Zhu *et al*. 2017) and subsequently subjected to Plant-CARE program (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot *et al*. 2002) for prediction of potential *cis*-acting elements.

Abiotic stress treatments

Seeds of upland cotton (*G. hirsutum* cv YZ1) were soaked in water and put at 28°C for one night. Next day drained out the excess moisture and put back at 28°C in incubators to induce germination, seeds with good growth potential were selected and planted into soil pots filled with commercially available sterilized composted soil under controlled conditions in growth chamber. The temperature, humidity and photoperiod cycle was set as 28°C, 60%, 16 h light/8 h dark period, respectively. Uniform cotton seedlings at 3–4 leaf stage were selected and subsequently subjected to salt (200 mmol/L NaCl) and heat stress (38°C) for 48 h. Alternatively, control seedlings were watered normally and used as mock. Leaf samples were collected after 72 h from both control and salt treated plants, immediately kept in liquid nitrogen and stored at -80°C for subsequent RNA extraction.

Expression pattern analysis

The plant total RNA from all samples was extracted using the method as reported earlier by Tu *et al.* (2007). After dilution, RNA was reverse transcribed into cDNA using the reverse transcription kit (Promega, U.S.A.). qRT-PCR was run as reported earlier by Xu *et al*. (2014) with thermal cycles was as follows, 95°C for 15 s, 60°C for 1 min and 95°C for 15 s. The $2^{-\Delta\Delta Ct}$ method was adopted to compare the gene expression values and *GhUBQ7* (DQ116441) gene (Xu *et al*. 2014) served as an internal control. Primers used for expression analysis are enlisted in Table S10.

For expression pattern analysis of *GhGASA* genes, publically available transcriptomic data related to different tissues (roots, stems, leaves, petals, ovules, seeds) and under multiple stresses including cold, heat, PEG and salinity were retrieved from Cotton FGD (https://cottonfgd.org/) (Zhu *et al*. 2017). Afterwards, Genesis software (Version 1.0) was used to normalize expression values and generation of heatmap (Sturn *et al*. 2002).

Results

Identification and characterization of GASA genes in *Gossypium* spp.

The genome-wide identification of GASA gene family in four species of cotton (*G. hirsutum*, *G. barbadense*, *G. arboreum* and *G. raimondii*) was carried out through using the *Arabidopsis* GASA domain as query in BLASTP search against the corresponding genome of four cotton species. As a result, total 116 GASA genes were obtained from the studied four *Gossypium* species (Table S1). The numbers of putative GASA genes were as 40 in *G. hirsutum*, 32 in *G. barbadense*, 22 in *G. arboreum* and 22 in *G. raimondii*. These identified genes were further manually checked using

NCBI conserved domain, SMART and Pfam databases. All the genes possessed the putative GASA domain as reported in previous studies (Fan *et al*. 2017; Zhang and Wang 2017; Ahmad *et al*. 2020). Further, these genes were named based on their chromosomal locations. Multiple sequence alignment showed that all these GASA family genes harbored 12 conserved cysteine residues, a characteristic motif of GASA gene family (Fig. S2). This motif is responsible for maintaining stability and structure and of GASA proteins (Betz 1993; Darby and Creighton 1995). Additionally, protein properties including length, molecular weights and isoelectric points of all the predicted GASA family members were analyzed using ExPASy program (Table S2).

A circos plot was created between two diploid species (*G. arboreum*, *G. raimondii*) and subgenomes of allotetraploid specie (*G. hirsutum*) to visualize the chromosomal distribution and syntenic relationship of GASA family genes. As shown in Fig. 1, all the GASA genes of studied species are evenly distributed on chromosomes. As *GaGASA*, *GrGASA* and *GhGASA* genes were found to be distributed on 10 chromosomes of A, D, and At and Dt subgenomes, respectively (Table S3). The distribution was uneven, with chromosomes A05 and A09 harbored 6 and 4 *GhGASA* genes while chromosomes A06, D05, D07, D09 and A07, D03, D04, D06 possess 3 and 2 *GhGASA* genes, respectively. Other chromosomes A02, A03, A04, A10, A11, A12, D02, D10, D11, and D12 contained single *GhGASA* gene (Fig. S1). Additionally, 40 and 42 duplicated gene pairs were identified from At to A2 and Dt to D5, respectively and members of each duplicated gene pairs exhibited great similarity to each other (Table S4).

Additional duplication analysis among *GhGASA* genes found 14 duplicated gene pairs and all of these pair experienced segmental duplications (Table S5). Further, All the duplicated *GhGASA* genes experienced purifying selection except two duplicated genes pairs (*GhGASA10*-*GhGASA31* and *GhGASA20*-*GhGASA39*) that undergoes positive selection (Table S5).

Phylogenetic clustering and Structural properties of GASA genes in cotton

For estimation of phylogenetic relationship among members of GASA family, a phylogenetic tree was constructed using amino acids sequences of model plant Arabidopsis and four cotton species. According to phylogenetic tree, GASA genes of cotton and Arabidopsis were clustered into three groups (G1, G2 and G3) based on their homology and protein structures (Fig. 2). Group 3 contained highest number of GASA family members 51, while group 1 and 2 possess 41 and 39 GASA genes, respectively.

The evolutionary relationship based on genes structural diversity was considered an important component in the study of multigene families. To study the structural similarity or differences among putative *GhGASA* genes, an exon-intron map was constructed (Fig. 3A). Results revealed variation in number of exons among *GhGASA* genes that fluctuated from 2 to 5, with highest number of introns and exons was found in *GhGASA17* (introns: 4, exons: 5). Expectedly, the number and composition of exon-intron between closely related members was same within the same group. More variations in number of exons and introns were observed among group 3 members. All members of group 2 contained one intron and 2 exons, while all members of group 1 include 3 introns and 4 exons, with the exceptions of *GhGASA19* and *GhGASA38* that contained 2 introns and 3 exons (Fig. 3). To get some perceptible information about paralogous *GhGASA* genes in phylogenetic tree, we analyzed their exon-intron structures. Interestingly, most of the paralogous *GhGASA* genes harbored same number and orientation of exons-introns. For example, *GhGASA1*/*30* possessed three exons and two introns but some variations were observed in intron length. Probably, these relations have formed the size and structures of putative GASA genes in cotton.

To further explore the structures and features of *GhGASA* gene family, MEME server was accessed for finding distinctive or similar motifs in *GhGASA* genes. Ten distinct motifs of different length were found among 40 *GhGASA* genes (Fig. 3B) and annotated using SMART and Pfam servers. Only motif 1 was found to be the representative of GASA domain, while the functions of other motifs were unknown (Table S6). Normally, most of the closely related members harbored similar motifs within the same group, proposing their similar function. For example, *GhGASA7/12*, *GhGASA14/34*, *GhGASA16/36*, *GhGASA9/25* and *GhGASA3/26* shared similar motifs. In detail investigation found that only motif 3 was present in all *GhGASA* genes, while motifs 4, 6, 7, 8 were absent in all members of group 1 and 2. Moreover, some motifs were found to be specific to special *GhGASA* members, such as motif 10 was only found in *GhGASA1* and *GhGASA24*. However, it is not known whether these motifs confer unique functions to these *GhGASA* genes or not.

Promoter analysis of *GhGASA* genes

Cis-elements are considered to participate in controlling expression of genes. To explore the probable association of these *cis*-elements with the expression or functions of *GhGASA* genes, 1.5 kb promoter region of each *GhGASA* genes was extracted and analyzed using PLANTCARE website. A diversity of *cis*-elements responsive to growth, development, stress, light and phytohormones were identified. Among 40 *GhGASA* genes, light responsive elements were predominant (57%), followed by hormones responsive (19%), growth and development related (14%) and environmental stress responsive (10%) (Fig. 4A and Table S7). The pattern of *cis*-elements varied among *GhGASA* genes. Surprisingly, *GhGASA11* harbored only one site for auxin from hormones responsive elements but

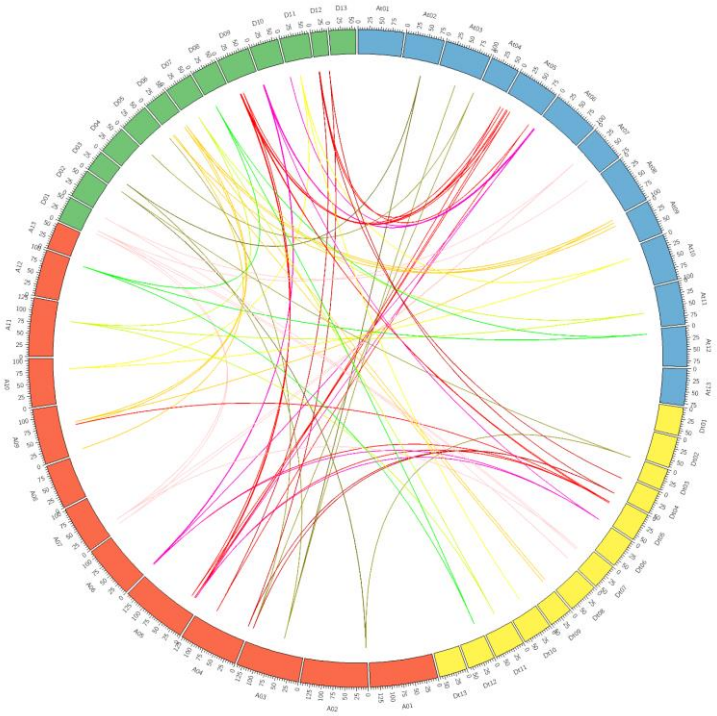

Fig. 1: Synteny analysis of GASA gene family
Syntenic association between two diploid species (*G. arboreum*, *G. raimondii*) and one tetraploid specie (*G. hirsutum*) was generated through circos program. Different chromosomes of selected species are marked with different colors

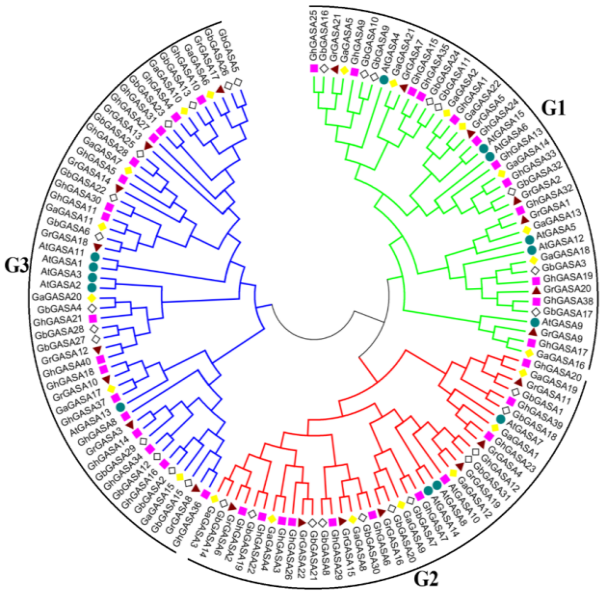

Fig. 2: Phylogenetic relationship of GASA genes from *Arabidopsis* and *Gossypium* species
Phylogenetic tree was constructed by MEGA 6.0 software using neighbor-joining (NJ) method with 1000 replicates. The GASA genes from *G. hirsutum*, *G. raimondii*, *G. arboreum*, *G. barbadense* and *Arabidopsis* were marked with different colors and shapes. Group1, Group 2 and Group 3 were indicated in green, red and blue colors, respectively

possess highest number of low temperature responsive *cis*-elements. Moreover, *GhGASA9* possessed highest number of gibberellin responsive sites among other *GhGASA* genes (Fig. 4B and Table S7).

Differential tissue-specific expression of *GhGASA* genes in upland cotton

To investigate the expression pattern of *GhGASA* genes in

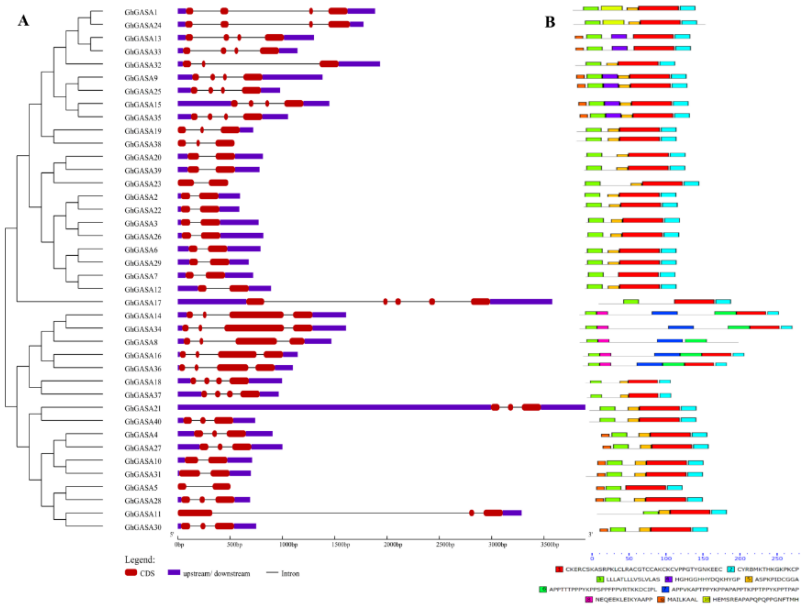

Fig. 3: Structure and motifs distribution of *GhGASA* genes

A. Gene structural analysis. Exons-introns structure of *GhGASA* genes. Exons, introns and upstream/downstream regions are shown by different colors. **B**. Motifs analysis. Protein motifs of *GhGASA* genes were represented by different colors and numbers

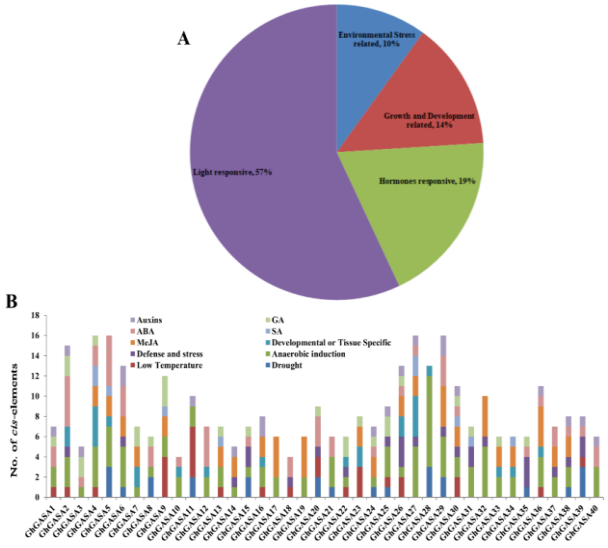

Fig. 4: Promoter region analysis of *GhGASA* genes

A. Percentage of *cis*-acting elements in the promoter regions of *GhGASA* genes, different colors corresponds to percentage of specific *cis*-elements category. **B**. presence of *cis*-elements in promoter region of each *GhGASA* gene was shown in column and in different colors

different tissues of cotton, a heatmap was generated. Results revealed that all the *GhGASA* genes were expressed in at least one tissue, except *GhGASA19* and *GhGASA29* (Fig. 5 and Table S8). However, three genes (*GhGASA17*, *GhGASA20*, and *GhGASA35*) expressed in all the seven tissues at all the time points [Fragments per kilobase of transcript per million mapped reads (FPKM ≥1)]. Moreover, four genes (*GhGASA10*, *GhGASA17*, *GhGASA20*, and *GhGASA35*) highly expressed in roots (FPKM ≥20) and six

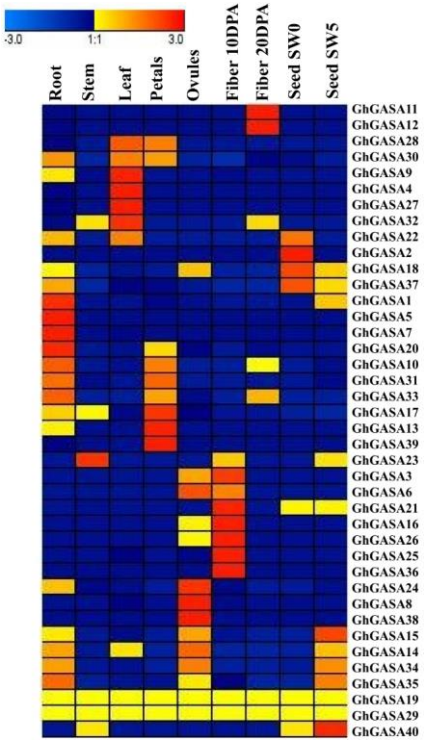

Fig. 5: Expression patterns of *GhGASA* genes in various tissues of upland cotton.

The RNA-seq data related to organ-specific expression were accessed from CottonFGD (https://cottonfgd.org/) and Genesis software package was used for generation of heatmaps. DPA (days post anthesis), SW0 (sowing in water 0 h), SW5 (sowing in water 5 h)

genes (*GhGASA4*, *GhGASA9*, *GhGASA27*, *GhGASA28*, *GhGASA30* and *GhGASA32*) were found to be dominantly

induced in leaves (FPKM ≥5). While, four genes (*GhGASA13*, *GhGASA17*, *GhGASA31* and *GhGASA39*) were highly induced in petals (FPKM ≥51) and five genes (*GhGASA6*, *GhGASA8*, *GhGASA14*, *GhGASA24*, and *GhGASA34* in ovules (FPKM ≥2.7). Expression of some genes found to be specific to single tissue, such as *GhGASA11*& *GhGASA12* only expressed in fiber (20 DPA).

Differential expression of *GhGASA* genes under multiple abiotic stresses

Considering important functions of GASA genes against various environmental stresses in a number of plant species, firstly, we thoroughly investigated the expression pattern of *GhGASA* genes using published transcriptomic data of cotton treated with heat, cold, polyethylene glycol (PEG) and salt (NaCl). From the results, it was observed that all *GhGASA* genes showed altered expression under one or more stress conditions except five *GhGASA* genes (*GhGASA5*, *GhGASA*19, *GhGASA*23, *GhGASA*29, *GhGASA*40) that were not expressed under any stress condition (Fig. 6 and Table S9). Comparing four treatments, more numbers of genes were differentially expressed against salt and cold stresses as compared with heat and PEG. In response to cold stress, seven *GhGASA* genes highly induced during 3 h time period and *GhGASA6/16* had the highest expression. Under heat stress, five genes showed increased expression at 12 h time period, with *GhGASA7* had the highest transcript abundance. Similarly, under salt stress seven *GhGASA* genes were highly expressed at 12 h time period (treatment RPKM/control RPKM ≥ 2.5) with highest expression noted for *GhGASA26*. Interestingly, some genes only expressed under specific stress condition, such as *GhGASA7/21* and *GhGASA6/16* only expressed in response to heat (12 h) and cold (3 h) (treatment RPKM/control RPKM ≥ 3). This specific expression pattern might support their involvement in modulating these stress responses in cotton. Secondly, to validate expression profile of *GhGASA* genes obtained through transcriptomic data, we choose eleven *GhGASA* genes based on their higher expression under salt or heat stress for further analysis through qRT-PCR. Similar to transcriptomic profile of *GhGASA* genes, most of the selected genes were up-regulated under salt stress as compared to heat stress. Four *GhGASA* genes including *GhGASA18*, *GhGASA22*, *GhGASA36*, and *GhGASA37* were highly expressed under salt stress. Expression of three genes including *GhGASA20*, *GhGASA33*, and *GhGASA39* was more under heat stress as compared to salt stress. However, three genes *GhGAS24*, *GhGASA30* and *GhGASA32* were down-regulated both under salt and heat stress.

Discussion

Earlier studies comprehensively showed the roles of GASA genes in monitoring plant developmental processes and responses to various stress responses in various plant species (Nahirnak *et al*. 2012b; Zhang and Wang 2017). However, previously it was not focused to identify and characterize GASA genes in cotton. In this study, a total of 116 putative GASA genes were identified from four species of cotton. Multiple sequence alignment of all these genes highlighted the existence of 12 amino acids conserved cysteine residues, a motif considered essential for maintaining structure and functions of GASA genes in different plant species (Betz 1993; Darby and Creighton 1995; Silverstein *et al*. 2007; Muhammad *et al*. 2019). In *G. hirsutum*, 40 GASA genes were identified, which is relatively a larger gene family among other three species of cotton (*G. barbadense*, *G. arboreum* and *G. raimondii*) and other reported plant species (Fan *et al*. 2017; Ahmad *et al*. 2019). Interestingly, all the *GaGASA*, *GrGASA*, and *GhGASA* are evenly distributed on 10 chromosomes of A, D and At and Dt subgenomes, respectively.

Gene duplication is an important mean for functional diversification, evolution and expansion of gene family (Kong *et al*. 2007). There are three ways through these duplication events takes in plants. Segmental duplications prevails when distribution of duplicated gene pairs are on different chromosomes, while tandem duplication occurs when duplicated genes are located on the same chromosomes. Duplication analysis among *GhGASA* genes revealed that segmental duplication is the leading cause for expansion of *GhGASA* gene family. Further, synonymous (Ks), non-synonymous (Ka) substitution rates and their ratio were computed to investigate the duplication mechanism of *GhGASA* genes after being diverged from their ancestor. The Ka/Ks ratio equal to 1 shows neutral selection, while greater than 1 and less than 1 indicates positive and purifying selection, respectively (Hurst *et al*. 2002). Interestingly, in our study most of the duplicated *GhGASA* genes experienced purifying selection mechanism. Additional systemic analysis is needed to explore further insights in the evolution of cotton GASA gene family.

In phylogenetic tree, all the GASA genes were distributed in three groups (G1, G2 and G3) based on their homology and structures. G3 contained more number of GASA genes than others. Similar kind of grouping among GASA family genes were also observed in other reported plant species (Zimmermann *et al*. 2010). Gene structure analysis showed that most of *GhGASA* genes within specific group harbored similar exons-introns number and organization corresponding to their conserved functionality. However, some variations were observed among group 3 members that indicate their functional diversity. In terms of exons-introns numbers, G2 was found to be more conserved than others, suggesting that other groups might gain or lost exons during evolutionary process leading to differences in structures. Similar trend of exons-introns conservation among G2 members was also noted in previous study (Ahmad *et al*. 2020). To further explore *GhGASA* genes in detail, we analyzed their motifs distribution. Expectedly,

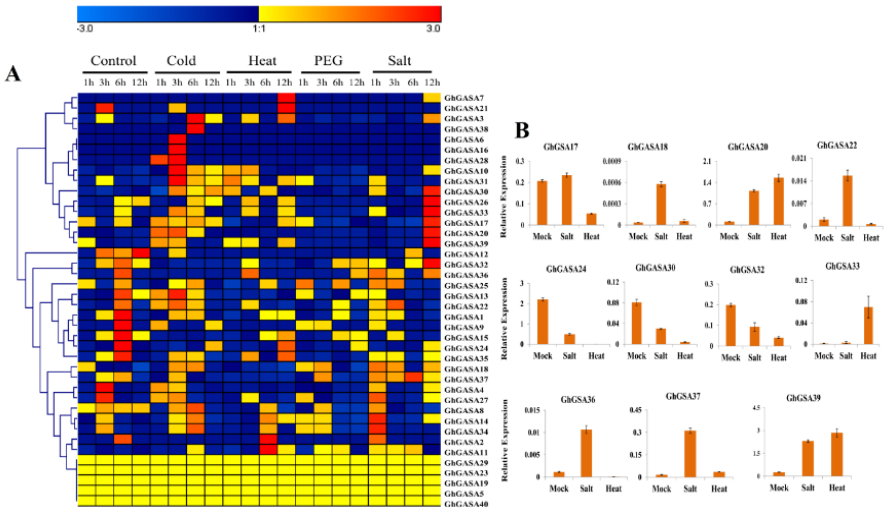

Fig. 6: Expression pattern of *GhGASA* genes against different abiotic stresses in upland cotton
A. Transcriptomic data related to heat, cold, PEG and drought were accessed from CottonFGD website (https://cottonfgd.org/), normalization and visualization of these data were performed using Genesis software package. **B**. Expression pattern of selected *GhGASA* genes under salt and heat stress by qRT-PCR. Columns represent the mean of three biological repeats and error bars represents the standard deviation

most of closely related *GhGASA* genes with in specific groups constitute similar motifs. Moreover, some specific motifs were found to be associated with some special *GhGASA* genes. However, it is not known whether these motifs confer unique functions to these *GhGASA* genes or not. In any case, conservation of these motifs within and between groups further supports the results of phylogenetic tree. *Cis*-elements present in gene promoter normally regulate gene expression and confers unique functionality to genes (Biłas *et al*. 2016). To study important roles of *GhGASA* genes under changing environmental conditions and various physiological processes of plants, we thoroughly investigated the promoter regions of each *GhGASA* gene. Mainly four types of elements were observed including growth, development-related, light, hormones and environmental stress-responsive elements. The diversity of *cis*-elements in *GhGASA* genes agree with the reported multidimensional functions of GASA genes in different plants (Oliveira-Lima *et al*. 2017; An *et al*. 2018; Li *et al*. 2019; Muhammad *et al*. 2019).

Expression pattern of genes provide useful clues for their functional characterization. Previously, it has been noted that GASA genes have spatiotemporal specificity in various plant species which might be due their probable involvement in diverse hormone signaling pathways (Wang *et al*. 2009; Zhang and Wang 2017). In present study, expression analysis of *GhGASA* genes in various tissues of cotton showed that maximum number of *GhGASA* genes induced in roots and ovules, signifying their important roles in development of these tissues. Moreover, considering the transcript abundance of *GhGASA* genes in cotton and the published role of their orthologs in *Arabidopsis* helped us to predict their probable functions. For example, *GhGASA35* along with higher expression in roots also strongly induced in seeds (SW5) (Fig. 5; Table S9) and its artholog *AtGASA4* in *Arabidopsis* regulates seed germination (Roxrud *et al*. 2007; Rubinovich and Weiss 2010). Similarly, *GhGASA2* did not expressed in any tissue except seed (SW0) and its orthologous gene *AtGASA10* reported to have important roles in seed germinations (Trapalis *et al*. 2017). This suggests that might be *GhGASA35* and *GhGASA2* have related roles in cotton as *AtGASA4* and *AtGASA10* in *Arabidopsis*. Nevertheless, this trend was not consistent to all *GhGASA* genes, signifying that these GASA genes might be undergone functional diversification across species.

The quality and yield of cotton is substantially affected by multiple abiotic stresses during the development of plant (Hassan *et al*. 2020). Therefore, we comprehensively studied the transcript abundance of *GhGASA* genes under several environmental cues including heat, cold, PEG and salt stress using published transcriptomic data. The results showed that most of *GhGASA* genes changed their expression under one or more stress conditions. Additionally, more numbers of *GhGASA* genes were highly induced under cold and salt stresses as compared with heat and PEG stresses, suggesting their probable function in monitoring these stresses in cotton. Further, qRT-PCR analysis of selected *GhGASA* genes under salt and heat stress supported the results of transcriptomic profile. Moreover, by comparing *GhGASA* genes in response to stress conditions and the existence of relevant *cis*-elements in promoter regions of those genes further support their effectiveness against these stresses. For example, *GhGASA26* was highly induced under salt stress (12 h) and contain maximum number of stress responsive elements in its promoter region. Interestingly, some *GhGASA* genes only induced under specific stress treatment, such as *GhGASA7/21* and *GhGASA6/16* only expressed in response

to heat (12 h) and cold (3 h) (treatment RPKM/control RPKM ≥ 3). This specific expression of selective genes might further support their importance in regulating these stress responses.

Conclusion

In silico analysis of cotton genomes enabled us to investigate GASA family genes in details. By adopting systematic approach consisting of conserved domain analysis, chromosomal distribution patterns, synteny, phylogeny, exons-introns organization and motifs division analysis, comprehensive characteristic features of GASA genes in cotton were elucidated. Promoter region analysis of *GhGASA* genes supported their involvement in a variety of biological functions in plants. Moreover, spatiotemporal tissue specific expression of *GhGASA* in upland cotton showed that most of *GhGASA* genes were induced in leaves and ovules than other. Additionally, qRT-PCR and transcriptomic expression profiles of *GhGASA* genes under various abiotic stress factors suggested that most of *GhGASA* genes have the capacity to regulate tolerance against multiple abiotic stresses. In short, the present genome-wide investigation of *GhGASA* gene family provides potential information for future more focused studies regarding in-depth characterization of GASA genes in cotton.

Acknowledgement

This study is a part of Interim Placement of Fresh Ph.D. program granted by Higher Education Commission of Pakistan.

Author Contributions

MS designed and wrote manuscript. AK, EN, MS, UA, AR helped in performing experiments and analyzing data. WM and MT helped in revising manuscript. AQ gave suggestions to improve the experimental work.

References

Ahmad B, J Yao, S Zhang, X Li, X Zhang, V Yadav, X Wang (2020). Genome-wide characterization and expression profiling of GASA genes during different stages of seed development in Grapevine (*Vitis vinifera* L.) predict their involvement in seed development. *Intl J Mol Sci* 21:1088–1103

Ahmad MZ, A Sana, A Jamil, JA Nasir, S Ahmed, MU Hameed, Abdullah (2019). A genome-wide approach to the comprehensive analysis of GASA gene family in *Glycine max*. *Plant Mol Biol* 100:607–620

Almasia NI, AA Bazzini, HE Hopp, C Vazquez-Rovere (2008). Overexpression of snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. *Mol Plant Pathol* 9:329–338

Alonso-Ramirez A, D Rodriguez, D Reyes, JA Jimenez, G Nicolas, M Lopez-Climent, A Gomez-Cadenas, C Nicolas (2009). Evidence for a role of gibberellins in salicylic acid-modulated early plant responses to abiotic stress in *Arabidopsis* seeds. *Plant Physiol* 150:1335–1344

An B, QN Wang, XD Zhang, B Zhang, HL Luo, CZ He (2018). Comprehensive transcriptional and functional analyses of HbGASA genes reveal their roles in fungal pathogen resistance in *Hevea brasiliensis*. *Tree Genet Genomics* 14; Article 41

Aubert D, M Chevillard, AM Dorne, G Arlaud, M Herzog (1998). Expression patterns of GASA genes in *Arabidopsis thaliana*: The GASA4 gene is up-regulated by gibberellins in meristematic regions. *Plant Mol Biol* 36:871–883

Bailey TL, J Johnson, CE Grant, WS Noble (2015). The MEME Suite. *Nucl Acids Res* 43:39–49

Balaji V, CD Smart (2012). Over-expression of snakin-2 and extensin-like protein genes restricts pathogen invasiveness and enhances tolerance to *Clavibacter michiganensis* subspp. michiganensis in transgenic tomato (*Solanum lycopersicum*). *Trans Res* 21:23–37

Ben-Nissan G, D Weiss (1996). The petunia homologue of tomato *gast1*: Transcript accumulation coincides with gibberellin-induced corolla cell elongation. *Plant Mol Biol* 32:1067–1074

Ben-Nissan G, JY Lee, A Borohov, D Weiss (2004). GIP, a *Petunia hybrida* GA-induced cysteine-rich protein: A possible role in shoot elongation and transition to flowering. *Plant J* 37:229–238

Berrocal-Lobo M, A Segura, M Moreno, G Lopez, F Garcia-Olmedo, A Molina (2002). Snakin-2, an antimicrobial peptide from potato whose gene is locally induced by wounding and responds to pathogen infection. *Plant Physiol* 128:951–961

Betz SF (1993). Disulfide bonds and the stability of globular proteins. *Protein Sci* 2:1551–1558

Biłas R, K Szafran, K Hnatuszko-Konka, AK Kononowicz (2016). *Cis*-regulatory elements used to control gene expression in plants. *Plant Cell Tiss Org Cult* 127:269–287

Bjellqvist B, B Basse, E Olsen, JE Celis (1994). Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. *Electrophoresis* 15:529–539

Ceserani T, A Trofka, N Gandotra, T Nelson (2009). VH1/BRL2 receptor-like kinase interacts with vascular-specific adaptor proteins VIT and VIK to influence leaf venation. *Plant J* 57:1000–1014

Dabbert T, MAJ Gore (2014). Challenges and perspectives on improving heat and drought stress resilience in cotton. *J Cotton Sci* 18:393–409

Darby N, TE Creighton (1995) Disulfide bonds in protein folding and stability. *In: Protein Stability and Folding*, pp:219–252. Springer, Dordrescht, The Netherlands

El-Gebali S, J Mistry, A Bateman, SR Eddy, A Luciani, SC Potter, M Qureshi, LJ Richardson, GA Salazar, A Smart, ELL Sonnhammer, L Hirsh, L Paladin, D Piovesan, SCE Tosatto, RD Finn (2019). The Pfam protein families database in 2019. The Pfam protein families database in 2019. *Nucl Acids Res* 47:427–432

Fan S, D Zhang, L Zhang, C Gao, M Xin, MM Tahir, Y Li, J Ma, M Han (2017). Comprehensive analysis of GASA family members in the *Malus domestica* genome: Identification, characterization, and their expressions in response to apple flower induction. *BMC Genomics* 18; Article 827

Fuente JIDL, I Amaya, C Castillejo, JF Sanchez-Sevilla, MA Quesada, MA Botella, V Valpuesta (2006). The strawberry gene FaGAST affects plant growth through inhibition of cell elongation. *J Exp Bot* 57:2401–2411

Furukawa T, N Sakaguchi, H Shimada (2006). Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. *Genes Genet Syst* 81:171–180

Haruta M, G Sabat, K Stecker, BB Minkoff, MR Sussman (2014). A peptide hormone and its receptor protein kinase regulate plant cell expansion. *Science* 343:408–411

Hassan A, M Ijaz, A Sattar, A Sher, I Rasheed, MZ Saleem, I Hussain (2020). Abiotic Stress Tolerance in Cotton. *In: Advances in Cotton Research*. IntechOpen, London, UK

Herzog M, AM Dorne, F Grellet (1995). GASA, a gibberellin-regulated gene family from *Arabidopsis thaliana* related to the tomato GAST1 gene. *Plant Mol Biol* 27:743–752

Hu B, J Jin, AY Guo, H Zhang, J Luo, G Gao (2015). GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* 31:1296–1297

Hurst LD, EJ Williams, C Pal (2002). Natural selection promotes the conservation of linkage of co-expressed genes. *Trends Genet* 18:604–606

Ko CB, YM Woo, DJ Lee, MC Lee, CS Kim (2007). Enhanced tolerance to heat stress in transgenic plants expressing the GASA4 gene. *Plant Physiol Biochem* 45:722–728

Kong H, LL Landherr, MW Frohlich, J Leebens-Mack, H Ma, CW dePamphilis (2007). Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. *Plant J* 50:873–885

Kotilainen M, Y Helariutta, M Mehto, E Pollanen, VA Albert, P Elomaa, TH Teeri (1999). GEG participates in the regulation of cell and organ shape during corolla and carpel development in *Gerbera hybrida*. *Plant Cell* 11:1093–1104

Kovalskaya N, RW Hammond (2009). Expression and functional characterization of the plant antimicrobial snakin-1 and defensin recombinant proteins. *Protein Expr Purif* 63:12–17

Lescot M, P Dehais, G Thijs, K Marchal, Y Moreau, YVD Peer, P Rouze, S Rombauts (2002). PlantCARE, a database of plant *cis*-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucl Acids Res* 30:325–327

Letunic I, T Doerks, P Bork (2015). SMART: recent updates, new developments and status in 2015. *Nucl Acids Res* 43:257–260

Li X, S Shi, Q Tao, Y Tao, J Miao, X Peng, C Li, Z Yang, Y Zhou, G Liang (2019). OsGASR9 positively regulates grain size and yield in rice (*Oryza sativa*). *Plant Sci* 286:17–27

Mao ZC, JY Zheng, YS Wang, GH Chen, YH Yang, DX Feng, BY Xie (2011). The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. *Phytoparasitica* 39:151–164

Marchler-Bauer A, MK Derbyshire, NR Gonzales, S Lu, F Chitsaz, LY Geer, RC Geer, J He, M Gwadz, DI Hurwitz, CJ Lanczycki, F Lu, GH Marchler, JS Song, N Thanki, Z Wang, RA Yamashita, D Zhang, C Zheng, SH Bryant (2015). CDD: NCBI's conserved domain database. *Nucl Acids Res* 43:222–226

Moyano-Canete E, ML Bellido, N Garcia-Caparros, L Medina-Puche, F Amil-Ruiz, JA Gonzalez-Reyes, JL Caballero, J Munoz-Blanco, R Blanco-Portales (2013). FaGAST2, a strawberry ripening-related gene, acts together with FaGAST1 to determine cell size of the fruit receptacle. *Plant Cell Physiol* 54:218–236

Muhammad I, WQ Li, XQ Jing, MR Zhou, A Shalmani, M Ali, XY Wei, R Sharif, WT Liu, KM Chen (2019). A systematic in silico prediction of gibberellic acid stimulated GASA family members: A novel small peptide contributes to floral architecture and transcriptomic changes induced by external stimuli in rice. *J Plant Physiol* 234–235:117–132

Nahirnak V, M Rivarola, MGD Urreta, N Paniego, HE Hopp, NI Almasia, C Vazquez-Rovere (2016). Genome-wide analysis of the Snakin/GASA gene family in *Solanum tuberosum* cv. Kennebec. *Amer J Potato Res* 93:172–188

Nahirnak V, NI Almasia, PV Fernandez, HE Hopp, JM Estevez, F Carrari, C Vazquez-Rovere (2012a). Potato Snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. *Plant Physiol* 158:252–263

Nahirnak V, NI Almasia, HE Hopp, C Vazquez-Rovere (2012b). Snakin/GASA proteins: involvement in hormone crosstalk and redox homeostasis. *Plant Signal Behav* 7:1004–1008

Oliveira-Lima M, AM Benko-Iseppon, J Neto, S Rodriguez-Decuadro, EA Kido, S Crovella, V Pandolfi (2017). Snakin: Structure, roles and applications of a plant antimicrobial peptide. *Curr Protein Pept Sci* 18:368–374

Peng J, L Lai, X Wang (2010). Temporal and spatial expression analysis of PRGL in *Gerbera hybrida*. *Mol Biol Rep* 37:3311–3317

Porto WF, OL Franco (2013). Theoretical structural insights into the snakin/GASA family. *Peptides* 44:163–167

Roxrud I, SE Lid, JC Fletcher, ED Schmidt, HG Opsahl-Sorteberg (2007). GASA4, one of the 14-member *Arabidopsis* GASA family of small polypeptides, regulates flowering and seed development. *Plant Cell Physiol* 48:471–483

Rubinovich L, D Weiss (2010). The *Arabidopsis* cysteine-rich protein GASA4 promotes GA responses and exhibits redox activity in bacteria and in planta. *Plant J* 64:1018–1027

Rubinovich L, S Ruthstein, D Weiss (2014). The *Arabidopsis* Cysteine-Rich GASA5 is a redox-active metalloprotein that suppresses Gibberellin responses. *Mol Plant* 7:244–247

Segura A, M Moreno, F Madueno, A Molina, F Garcia-Olmedo (1999). Snakin-1, a peptide from potato that is active against plant pathogens. *Mol Plant Microb* 12:16–23

Shi L, RT Gast, M Gopalraj, NE Olszewski (1992). Characterization of a shoot-specific, GA3- and ABA-regulated gene from tomato. *Plant J* 2:153–159

Silverstein KA, WA Moskal, HC Wu, BA Underwood, MA Graham, CD Town, KA VandenBosch (2007). Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. *Plant J* 51:262–280

Sturn A, J Quackenbush, Z Trajanoski (2002). Genesis: cluster analysis of microarray data. *Bioinformatics* 18:207–208

Sun S, H Wang, H Yu, C Zhong, X Zhang, J Peng, X Wang (2013). GASA14 regulates leaf expansion and abiotic stress resistance by modulating reactive oxygen species accumulation. *J Exp Bot* 64:1637–1647

Suyama M, D Torrents, P Bork (2006). PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucl Acids Res* 34:609–612

Tamura K, G Stecher, D Peterson, A Filipski, S Kumar (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* 30:2725–2729

Thompson JD, TJ Gibson, DG Higgins (2002). Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinform* 1:2-3

Trapalis M, SF Li, RW Parish (2017). The *Arabidopsis* GASA10 gene encodes a cell wall protein strongly expressed in developing anthers and seeds. *Plant Sci* 260:71–79

Tu LL, XL Zhang, SG Liang, DQ Liu, LF Zhu, FC Zeng, YC Nie, XP Guo, FL Deng, JF Tan, L Xu (2007). Genes expression analyses of sea-island cotton (*Gossypium barbadense* L.) during fiber development. *Plant Cell Rep* 26:1309–1320

Voorrips RE (2002). MapChart: Software for the graphical presentation of linkage maps and QTLs. *J Hered* 93:77–78

Wang L, Z Wang, YY Xu, SH Joo, SK Kim, Z Xue, ZH Xu, ZY Wang, K Chong (2009). OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice. *Plant J* 57:498–510

Wigoda N, G Ben-Nissan, D Granot, A Schwartz, D Weiss (2006). The gibberellin-induced, cysteine-rich protein GIP2 from *Petunia hybrida* exhibits in planta antioxidant activity. *Plant J* 48:796–805

Xu L, W Zhang, X He, M Liu, K Zhang, M Shaban, L Sun, J Zhu, Y Luo, D Yuan, X Zhang, L Zhu (2014). Functional characterization of cotton genes responsive to *Verticillium dahliae* through bioinformatics and reverse genetics strategies. *J Exp Bot* 65:6679–6692

Yang S, X Zhang, JX Yue, D Tian, JQ Chen (2008). Recent duplications dominate NBS-encoding gene expansion in two woody species. *Mol Genet Genom* 280:187–198

Zhang LY, XL Geng, HY Zhang, CL Zhou, AJ Zhao, F Wang, Y Zhao, XJ Tian, ZR Hu, MM Xin, YY Yao, ZF Ni, QX Sun, HR Peng (2017). Isolation and characterization of heat-responsive gene TaGASR1 from wheat (*Triticum aestivum* L.). *J Plant Biol* 60:57–65

Zhang S, X Wang (2017). One new kind of phytohormonal signaling integrator: Up-and-coming GASA family genes. *Plant Signal Behav* 12:1-7

Zhang S, X Wang (2011). Overexpression of GASA5 increases the sensitivity of *Arabidopsis* to heat stress. *J Plant Physiol* 168:2093–2101

Zhang S, C Yang, J Peng, S Sun, X Wang (2009). GASA5, a regulator of flowering time and stem growth in *Arabidopsis thaliana*. *Plant Mol Biol* 69:745–759

Zhu T, C Liang, Z Meng, G Sun, Z Meng, S Guo, R Zhang (2017). CottonFGD: an integrated functional genomics database for cotton. *BMC Plant Biol* 17:101-109

Zimmermann R, H Sakai, F Hochholdinger (2010). The Gibberellic Acid Stimulated-Like gene family in maize and its role in lateral root development. *Plant Physiol* 152:356–365