BMC Genomics

RESEARCH

Open Access

Comprehensive identification of *GASA* genes in sunflower and expression profiling in response to drought

Muhammad Asad Ullah[1], Muhammad Awais Ahmed[1], Latifa AlHusnain[2], Muhammad Abu Bakar Zia[3], Muneera D. F. AlKahtani[2], Kotb A. Attia[4] and Mohammed Hawash[5*]

Abstract

Drought stress poses a critical threat to global crop yields and sustainable agriculture. The *GASA* genes are recognized for their pivotal role in stress tolerance and plant growth, but little is known about how they function in sunflowers. The investigation aimed to identify and elucidate the role of *HaGASA* genes in conferring sunflowers with drought tolerance. Twenty-seven different *HaGASA* gene family members were found in this study that were inconsistently located across eleven sunflower chromosomes. Phylogeny analysis revealed that the sunflower *HaGASA* genes were divided into five subgroups by comparing *GASA* genes with those from Arabidopsis, peanut, and soybean, with members within each subgroup displaying similar conserved motifs and gene structures. In-silico evaluation of cis-regulatory elements indicated the existence of specific elements associated with stress-responsiveness being the most abundant, followed by hormone, light, and growth-responsive elements. Transcriptomic data from the NCBI database was utilized to assess the *HaGASA* genes expression profile in different sunflower varieties under drought conditions. The *HaGASA* genes expression across ten sunflower genotypes under drought stress, revealed 14 differentially expressed *HaGASA* genes, implying their active role in the plant's stress response. The expression in different organs revealed that *HaGASA2*, *HaGASA11*, *HaGASA17*, *HaGASA19*, *HaGASA21* and *HaGASA26* displayed maximum expression in the stem. Our findings implicate *HaGASA* genes in mediating sunflower growth maintenance and adaptation to abiotic stress, particularly drought. The findings, taken together, provided a basic understanding of the structure and potential functions of *HaGASA* genes, setting the framework for further functional investigations into their roles in drought stress mitigation and crop improvement strategies.

Keywords *GASA*, Evolution, Expression analysis, Sunflower, Drought

*Correspondence:
Mohammed Hawash
mohawash@najah.edu
[1] Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab, P.O BOX. 54590, Lahore, Pakistan
[2] Department of Biology, College of Science, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia
[3] Department of Plant Breeding and Genetics, Faculty of Agriculture Sciences and Technology, University of Layyah, P.O BOX 31200, Layyah, Pakistan
[4] Center of Excellence in Biotechnology Research, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
[5] Department of Pharmacy, Faculty of Medicine and Health Sciences, An-Najah National University, P.O. Box 7, Nablus, Palestine

Introduction

Plants rely on a complex framework of genes to regulate various aspects of growth, development, and stress responses. This adaptability is partly attributed to small, cysteine-rich proteins. Cysteine-rich-peptide (CRP) is a class of proteins consisting of eight sub-families namely: Snakins, defensins, thionins, hevein-like peptides, nonspecific lipid transfer proteins (LTPs), knottins, α-hairpinins, and cyclic peptide [1]. Several new classes of Cysteine-rich-peptide CRPs have been identified, which expands our understanding of their crucial roles in plant biology. These include impatiens balsamina Antimicrobial Peptides (*Ib-AMPs*) known for their potent antibacterial properties against pathogens like Methicillin-resistant Staphylococcus aureus (*MRSA*) [2]. Additionally Rapid Alkalinization Factor (*RALF*) play a crucial role in plant development and stress responses, regulating growth, development, and stress adaptation through receptor kinases and intercellular communication [3]. Epidermal Patterning Factor (*EPF*) and its associated *EPF*-Like (*EPFL*) are necessary for stomatal growth and patterning with conserved cysteine residues contributing to their stability and functioning [3–5]. Furthermore Maize Egg Apparatus (*EA1*) play a vital role in facilitating pollen tube guidance towards the ovule [6]. Recent research has also identified at least 23 cysteine-rich peptides in *Citrullus colocynthis*, with eight novel peptides named citcol-1 to citcol-8 characterized for their structural features and potential bioactivities, including antimicrobial functions [7].

CRP proteins represent a substantial family of genes, specifically classified under the *GASA* (Gibberellic Acid-Stimulated in Arabidopsis), *GAST* (Gibberellic Acid-Stimulated Transcripts), *GASR* (Gibberellic Acid-Stimulated Regulators), and snakin subfamilies. These proteins are characterized by their relatively short amino acid sequences and low molecular weights, playing critical roles in various physiological processes [8–11] that are mostly controlled by gibberellins [12, 13]. The *GASA* domain is a highly conserved protein with three regions including an N-terminal that has a signal region of 18–29AA, an extremely varying, hydrophilic region of 7–31 AA at the middle part; and a C-terminus area of 60 AA consisting of 12 residues of cysteine that helps facilitate molecule's biochemical stability [14]. The N-terminal part is essential to interact with other proteins while the C-terminal likely plays a role in DNA binding site and transcriptional regulation [15]. The *GASA* gene was first identified in 1992 with the discovery of its initial member GAST1 in tomato [10]. Moreover, many researchers have also characterized *GASA* homologs in Arabidopsis (15 genes) [14], Rice (10 genes) [16], Apple (26 genes) [17], Cucumber (09 genes) [18], Cotton (38 genes) [19], Peanut (40 genes) [20], Wheat (37 genes) [21], Soybean (37 genes) [22], Sorghum (12 genes) [23], Chinese cabbage (15 genes) [24], Potato (16 genes) [25], Citrus (18 genes) [26], Pine apple (15 genes) [27], Grapevine (14genes) [9] tobacco (18 genes) [28], *Populus trichocarpa* (15 genes) [29], Common bean (23 genes) [30], Tomato (17 genes) [31], Maize (10 genes) [32] and Strawberry (2 genes) [33].

The *GASA* proteins have been known to govern various aspects of plant development and hormone regulation. The *AtGASA4* has been found to be involved in the light signaling pathway and promotes flowering time in the Arabidopsis, a model plant [14]. Furthermore, research on rice reveals the gene *OsGSR1* stimulates BR, (brassinosteroids) production by actively regulating DIM/DWF1, a BR biosynthetic enzyme that converts 24-methylenecholesterol to campesterol [34]. A transgenic Arabidopsis plant overexpressing a gibberellin-responsive gene from beechnut reduced GA dependence for growth and improved seed germination and establishment responses to salinity, oxidative, and heat stress [35]. The gene *GsGASA1* is implicated in the inhibition of root growth under chronic cold stress, a process mediated by the accumulation of DELLA proteins [36]. The *OsGASR9* is potential in regulating grain size and yield via the GA pathway [37]. *GASA* proteins also confer resilience to various abiotic and biotic stresses in plants. For instance, *SmGASA4* enhances plant resistance to drought, salinity, and paclobutrazol (PBZ) stress [38]. The *SN1* gene serves as a defense mechanism against *C. michiganensis subsp. sepedonicus* [39]. A novel *CaSN* gene from the sankins family confers resistance in pepper against root-knot nematode infection [40].

Sunflower (*Helianthus annuus L.*) are economically most important oilseed crop grown throughout the world for their edible seeds and oil [41]. The extreme conditions such as drought, high salinity, Heat stress, and heavy metal stress in different crops substantially influence crop yield and quality [42, 43]. Drought is a key abiotic stress that greatly affects crop development and output by altering critical metabolic pathways and influencing various physiological and biochemical features [44]. Various genes in sunflower including the *SAP* gene family [45], *NAC* transcription family [46],*Dof* gene family [47], *MYB* gene [48], have been analyzed in response to drought. Still, the *GASA* gene known for its involvement in regulating drought remains uncharacterized in Sunflower. This knowledge gap hinders our ability to exploit the potential of *GASA* genes for sunflower improvement. Our study aims to address this knowledge gap by thoroughly analyzing the *GASA* gene family in sunflower under drought conditions. The *GASA* gene family in the sunflower is characterized on a genome-wide scale and

analyzed expression analysis under drought in this study. We find out the evolutionary relationship of *GASA* genes and investigate the functional differentiation in the sunflower genome. This research will provide a basis for future functional studies for the development of drought-tolerant varieties.

Materials and methods

Database hunting and retrieval of *HaGASA* sequences

The *GASA* domain query sequence was retrieved from the NCBI online database (https://www.ncbi.nlm.nih.gov/, 12 February 2024) using accession AAG23437.1 [19]. PFAM, an online database was utilized to acquire the Hidden Markov Model profile (PF02704) for the *GASA* domain [18]. The query sequence was subjected to a BLAST search against the sunflower genome in the Phytozome v.13 (https://phytozome-next.jgi.doe.gov/, 13 February 2024) database to identify potential *HaGASA* genes (https://phytozome-next.jgi.doe.gov/, 13 February 2024). Twenty-seven sunflower *GASA* genes were validated using Motif-Finder (https://www.genome.jp/tools/motif/, 13 February 2024) to validate the existence of the conserved *GASA* domain within each gene sequence [49].

Determining physical & chemical characteristics and cellular localization of *HaGASA* genes

The characteristics like gravy, isoelectric point, protein length, and weight were determined using were determined through a web tool Protopram [50]. Further gene positioning traits including amino acids, direction, and start and end points were retrieved through Phytozome v.13 [51].

Localization of *HaGASA* genes in different organelles was predicted through Wolf Psort, a web tool (https://wolfpsort.hgc.jp/, 14 February 2024) [52]. TBtools software was utilized for creating a heatmap to visually represent the location of each gene in different cellular organelles [53].

Analysis of evolutionary relations, gene structure and conserved motifs of *HaSASA* genes

The peptide sequences from *GASA* genes in Arabidopsis, peanut, soybean, and sunflower were aligned using Mega11 software. To achieve statistical reliability, a neighbor-joining evolutionary tree was constructed with a bootstrap value of 1000 [54]. The tree was exported to iTol for improving visual presentation [55].

Genomic and CDS sequences of *HaGASA* were retrieved through Phytozome. The intron-exon architecture of *HaGASA* genes was revealed through the Gene structure display serve (https://gsds.gao-lab.org/, 18 February 2024) using genomic and CDS sequences [56]. Conserved motifs analysis was performed using meme suit (https://meme-suite.org/meme/tools/meme, 18 February 2024) by analyzing 15 motifs [57]. The motifs were visually represented through an "advanced gene structure view" setup in TBtools using a conserved domain file, meme suit, and phylogeny file [58].

Chromosomal localization, duplication and collinearity analysis

Data on the location of The *HaGASA* genes positioning data was retrieved from the Phytozome. TBtools software was subsequently employed to build a graphic representation of the sunflower chromosomes, illustrating the precise genomic locations of the *HaGASA* genes [59]. Ka and Ks values were calculated using the ka-ks calculator function in TBtools [60].

The one-step MCScan function of TBtools was used to observe collinear relations by comparing the genomes of sunflower, peanut, soybean, and Arabidopsis [61]. A map showing syntenic relations was created using the advanced circos function in TBtools [62].

Prediction of Cis-Regulatory Elements (CREs)

1500 bps upstream sequence starting from the initiation codon was taken to analyze the promoter region of each *HaGASA* gene [63]. CREs present in these specific regions were predicted through PlantCare an online database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/, 3 March 2024) [64]. CREs were divided into different classes and a heatmap was created using TBtools [65].

HaGASA genes enrichment analysis

The *HaGASA* genes underwent ontology analysis using a web tool Shiny GO 0.80 (http://bioinformatics.sdstate.edu/go/, 7 March 2024) to estimate their potential involvement in specific biological processes and cellular components [66]. The enrichment level was calculated by applying a 0.01 *p* value [67].

Determination of microRNAs targeted sites

An online database PmiREN (http://www.pmiren.com/, 8 March 2024) was utilized to retrieve miRNA datasets of sunflower [68]. The CDS sequence of all *HaGASA* was searched using psRNA Target (https://plantgrn.noble.org/psRNATarget/analysis?function=3, 8 March 2024) to find miRNAs targeting *HaGASA* genes [69].

HaGASA genes transcriptomic analysis

The expression profile of *HaGASA* genes in sunflower varieties SF-01, SF-02, SF-03, SF-04, SF-05, SF-06, SF-07, SF-08, SF-09, and SF-10 were extracted from the publically accessible database NCBI GEO database

(https://www.ncbi.nlm.nih.gov/geo/, 2 April 2024) to investigate the drought stress response of the *GASA* gene family. Ten diverse sunflower genotypes underwent exposure to controlled drought stress conditions during the vegetative phase to evaluate the transcriptional responsiveness of *HaGASA* genes. Leaf tissues were excited from both control and water-stressed individuals of each genotype and used for transcriptome profiling to elucidate the expression patterns of *HaGASA* genes in response to drought stress [70]. Two-way ANOVA was applied to check the significant expression of *HaGASA* genes. Variation in the expression level of *HaGASA* genes was analyzed in five different tissues including stem, root, axil, leaf and flower [71]. To achieve statistical reliability, the experiment included three replicates. The RNA sequencing data from specific tissues were especially evaluated during the plant flowering phase.

Results

HaGASA genes identification, subcellular localization & physicochemical properties

To identify *GASA* genes in the sunflower, the sequence of the *GASA* domain was blasted against the sunflower genome using Phytozome. A total of 27 genes were identified that were further analyzed to remove non-redundant proteins (Table 1). All of the genes were confirmed to contain the *GASA* domain. The twenty-seven proteins identified as *GASA* in the sunflower consisted of *HaGASA1-HaGASA27*.

The *HaGASA* genes encoded proteins that ranged from 88 to 431 amino acids (AA), of which *HaGASA23* contained a maximum number of 431 AA. The molecular weights span from 9.6 to 27.5 kDa, with *HaGASA9* displaying the highest molecular mass. The *HaGASA* genes isoelectric point shown ranges from 6.6 to 9.6, all of these genes are considered to be unstable as the value is lower than 40. All *HaGASA* genes except *HaGASA1*, *HaGASA3*, and *HaGASA5* contained negative gravy values. *HaGASA* genes *HaGASA1*, *HaGASA3*, and *HaGASA5* are hydrophobic, while the others are hydrophilic. The GRAVY score is essential for determining the hydropathy of each protein and is required for a thorough investigation of physicochemical properties.

Subcellular localization of *HaGASA* genes was analyzed to predict the role of genes based on location. All *HaGASA* genes depicted extracellular localization followed by plasma membrane and nucleus. Wolfpsort was further utilized to predict the number of genes localized in various organelles based on the amino acid composition of each gene (Fig. 1).

Exploration of evolutionary relationships, structures, and comparative motifs of*HaGASA* genes

Phylogeny connections of *GASA* genes in the sunflower were revealed through Mega11 software [72]. A tree was constructed encompassing 120 *GASA* proteins from four different species (15 *AtGASA* genes from Arabidopsis, 37 *GmGASA* genes from soybean, 41 *AhGASA* genes from peanut, and 27 *HaGASA* genes from sunflower) (Fig. 2). Arabidopsis *GASA* genes were used as a reference to classify the tree, which is divided into five sub-groups. Group three was the largest, containing seven *HaGASA* genes, while group four was the smallest, with only four *HaGASA* genes. Specifically, the tree included eight *HaGASA* genes in the 1st group, four in the 2nd group, four in the 3rd group, four in the 4th group, and seven in the 5th group.

The evolutionary history of a gene family can be verified based on the architecture of exons and introns of that gene [73]. The pattern of exons and introns can be used as a mark for studying the evolution of a gene. The analysis revealed that exons and introns division depicted consistency with the evolutionary tree of *HaGASA* genes. The exon number of *HaGASA* genes varied from 2 to 4, and *HaGASA10* and *HaGASA25* contained maximum exons (Fig. 3).

The presence of conserved motifs can be further utilized to classify a gene family. Ten motifs of 15–50 amino acids were analyzed in each gene using MEME [74]. Motifs 1, 2, and 3 were universally present in all *HaGASA* genes, Motif 4 was found in all except *HaGASA18*, Motif 8 was exclusive to *HaGASA13*, *HaGASA21*, and *HaGASA25*, and Motif 9 was detected only in *HaGASA2* and *HaGASA21* (Fig. 4).

Chromosomal mapping of *HaGASA* genes

Chromosomal mapping revealed that all *HaGASA* genes were located on 11 different chromosomes of sunflower (Fig. 5). Chromosomes 2, 3, 5, 7, 8, 9, 10, 12, 14, 15 and 17 contained all *HaGASA* genes. The highest number of *HaGASA* genes (nine) were mapped to chromosome 14, with chromosomes 7 and 10 each harboring four genes, chromosome 9 containing three genes, and the remaining chromosomes each hosting a single gene.

Collinearity and duplication analysis of *HaGASA* genes

Multiple synteny plot was created to reveal the evolutionary relation of *HaGASA* genes with other species like Arabidopsis, peanut, and soybean (Fig. 6). Collinearity analysis revealed that multiple copies of *HaGASA* genes were present in other species, with eight copies found in the *Arachis hypogaea* genome, eight copies

Table 1 Comprehensive information about various characteristics of the HaGASA genes

Genes ID	Accession No.	Chr.No.	Chr. Location(bp)		Direction	Size(AA)		Molecular weight (kDa)	Pi value	Exon	Intron	GRAVY	Cellular localization
			Start	End		mRNA(CDS)	Peptide						
HaGASA1	HanXRQChr10g0286031	10	44,481,238	44,481,684	Reverse	282	93	9.87247	6.67	**3**	**2**	0.02	Extracellular
HaGASA2	HanXRQChr10g0294941	10	109,429,223	109,431,778	Reverse	357	118	13.18129	9.05	2	1	-0.277	Extracellular
HaGASA3	HanXRQChr10g0294961	10	109,644,207	109,645,625	Reverse	291	96	10.33219	8.77	3	2	0.07	Extracellular
HaGASA4	HanXRQChr10g0294901	10	109,361,221	109,364,395	Reverse	309	102	10.91776	9.13	3	2	-0.085	Extracellular
HaGASA5	HanXRQChr07g0201841	7	86,428,894	86,429,394	Forward	282	93	9.98464	7.44	3	2	0.044	Extracellular
HaGASA6	HanXRQChr07g0199351	7	78,968,935	78,970,584	Reverse	303	100	10.82881	9.3	3	2	-0.029	Extracellular
HaGASA7	HanXRQChr07g0189081	7	18,315,753	18,317,160	Forward	291	96	10.73392	9.55	3	2	-0.107	Extracellular
HaGASA8	HanXRQChr07g0188921	7	17,915,467	17,917,125	Forward	294	97	10.77884	9.5	3	2	-0.171	Extracellular
HaGASA9	HanXRQChr09g0259181	9	146,932,421	146,933,728	Forward	783	260	27.57226	8.8	3	2	-0.233	Plasma membrane, Nuclear
HaGASA10	HanXRQChr09g0266031	9	174,200,046	174,201,506	Forward	597	198	20.97079	9.67	4	3	-0.609	Plasma membrane, Nuclear
HaGASA11	HanXRQChr09g0266481	9	176,131,346	176,133,025	Reverse	273	90	9.76447	9.09	2	1	-0.189	Extracellular
HaGASA12	HanXRQChr05g0155161	5	188,074,999	188,075,532	Reverse	369	122	13.46478	8.2	3	2	0.052	Extracellular
HaGASA13	HanXRQChr12g0373281	12	77,118,105	77,118,778	Forward	396	131	14.65733	9.52	2	1	-0.318	Extracellular
HaGASA14	HanXRQChr14g0430041	14	25,110,087	25,112,756	Reverse	297	98	10.49436	8.9	3	2	-0.057	Extracellular
HaGASA15	HanXRQChr14g0449591	14	138,912,152	138,912,886	Reverse	324	107	11.35925	8.61	2	1	-0.09	Extracellular
HaGASA16	HanXRQChr14g0434931	14	66,170,595	66,171,047	Reverse	282	93	9.93257	7.43	3	2	0.037	Extracellular
HaGASA17	HanXRQChr14g0439241	14	92,732,536	92,733,258	Reverse	309	102	10.72855	8.86	2	1	-0.118	Extracellular
HaGASA18	HanXRQChr14g0433221	14	53,361,104	53,365,292	Forward	351	116	12.74614	9.1	4	3	-0.263	Extracellular
HaGASA19	HanXRQChr14g0449771	14	139,413,488	139,414,511	Reverse	302	100	10.88685	9.2	3	2	-0.047	Extracellular
HaGASA20	HanXRQChr14g0444731	14	120,372,650	120,373,869	Reverse	282	93	10.34129	9.27	3	2	-0.124	Extracellular
HaGASA21	HanXRQChr14g0437871	14	84,963,265	84,966,213	Forward	459	152	16.95315	9.18	3	2	-0.013	Extracellular
HaGASA22	HanXRQChr14g0439261	14	92,787,320	92,788,451	Reverse	300	99	10.57835	8.76	2	1	-0.198	Extracellular
HaGASA23	HanXRQChr17g0548361	17	5,192,246	55,194,197	Reverse	1296	431	43.90334	8.89	3	2	-0.659	Nuclear, Plasma membrane
HaGASA24	HanXRQChr03g0068801	3	57,299,486	57,300,688	Reverse	270	89	10.2014	9.13	2	1	-0.176	Extracellular
HaGASA25	HanXRQChr15g0476611	15	48,139,354	4,814,385	Forward	327	108	11.81776	8.93	4	3	-0.094	Extracellular
HaGASA26	HanXRQChr02g0039561	2	48,843,745	48,844,592	Reverse	270	89	9.71947	8.46	2	1	-0.061	Extracellular
HaGASA27	HanXRQChr08g0211931	8	12,006,346	12,007,025	Forward	267	88	9.6383	8.62	2	1	-0.131	Extracellular

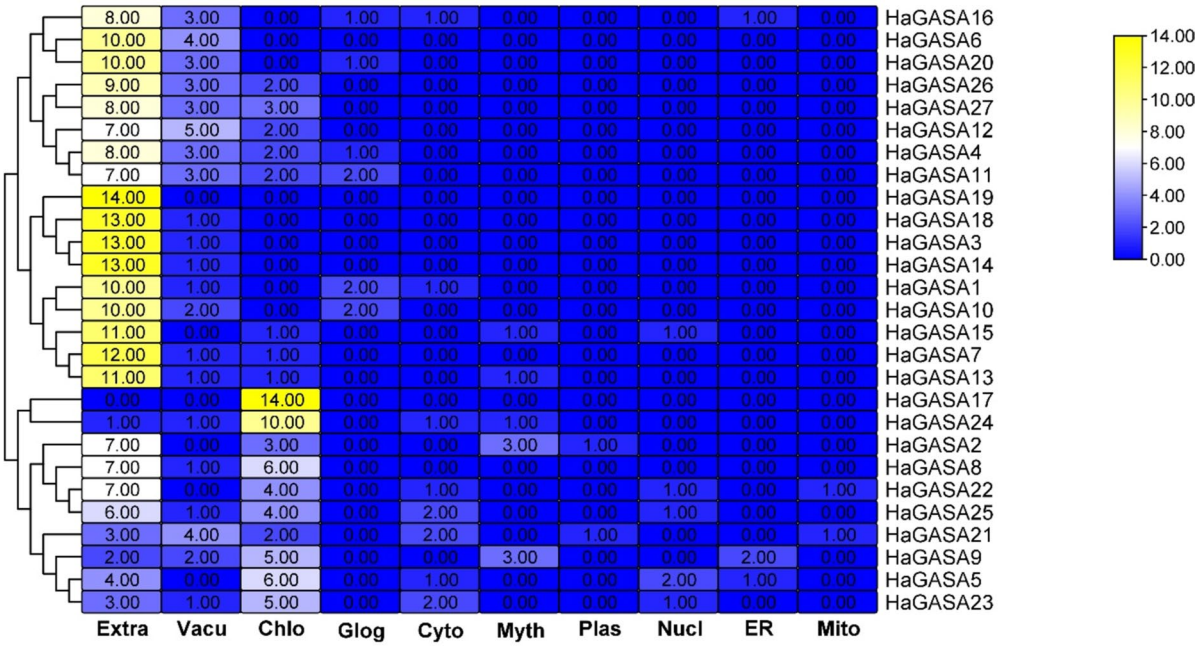

Fig. 1 Subcellular localization of HaGASA genes in different cellular organelles

in the *Glycine max* genome, and a single copy in the *Arabidopsis thaliana* genome.

The *HaGASA* gene duplication events were predicted by ka and ks estimation using TBtools. Where Ka counts the rate of non-syn substitutions per nonsynonymous site, Ks counts the proportion of synonymous substitutions per synonymous site [75, 76]. Selective pressure on *HaGASA* genes is indicated by the Ka/Ks ratio, which compares the number of non-syn to synonymous substitutions. The Ka/Ks value ranged from 0.06 (*HaGASA6_HaGASA19* pair) to 0.58 (*HaGASA2_HaGASA3* pair) across the eight paralogous pairs of *HaGASA* genes (Fig. 7). All paralogous *HaGASA* proteins in the sunflower had Ka/Ks ratios lower than 1. This data implies the probability of low functional convergence in the duplication process, possibly because of the predominance of purifying selection. The date of eight paralogous *HaGASA* genes varied from 3.8 (*HaGASA1-HaGASA16*) to 34.3 (*HaGASA13-HaGASA22*) million years ago. The *HaGASA13-HaGASA22* gene was considered to be the most primitive.

Gene enrichment analysis

The *HaGASA* gene functioning was further comprehensively understood using gene enrichment. The *HaGASA* genes were further classified into biological processes and cellular components (Fig. 8). In the biological process category, the genes were predominantly involved in various processes mostly gibberellic-acid related pathways. Most of the genes were enriched in extracellular regions based on the cellular component category. The analysis can be utilized to predict diverse functions of *HaGASA* genes in cellular metabolisms.

Analysis of CREs of *HaGASA* genes

Each gene's transcriptional expression is highly influenced by specific elements present in the promoter regions at the binding sites of that gene [77]. Therefore, these specific CREs were in-silico evaluated for speculating gene function [78]. The CREs present in *HaGASA* genes were divided into four categories based on specific functions. The *HaGASA* genes contained maximum stress-responsive CREs (53%) following hormone-responsive (23%), light-responsive (16%) and growth-responsive elements (8%) (Fig. 9). Ten CREs associated with stress responses, including LTR, ARE, GC-motif, DRE core, Box 4, MYB, MYC, STRE, WRE3, and W box, were identified in *HaGASA* genes, with MYC, MYB, and Box 4 being the most frequent. Within the hormone-responsive category, ten cis-regulatory elements (ABRE, AuxRR-core, ERE, GARE-motif, P-box, TCA, TCA-element, TCT-motif, TGA-element, and TGACG-motif) were analyzed, revealing that the ERE motif was the most prevalent. Thirteen elements (ACE, AE-box, ATCT-motif, AT1-motif, 3-AF1 binding site, Box II, G-box, GA-motif, GATA-motif, GT1-motif, I-box, Gap-box, and MRE) have been analyzed in the light-responsive category, with the G-box and GT1-motif emerging as the most abundant. Eight elements (AAGAA-motif, CCG

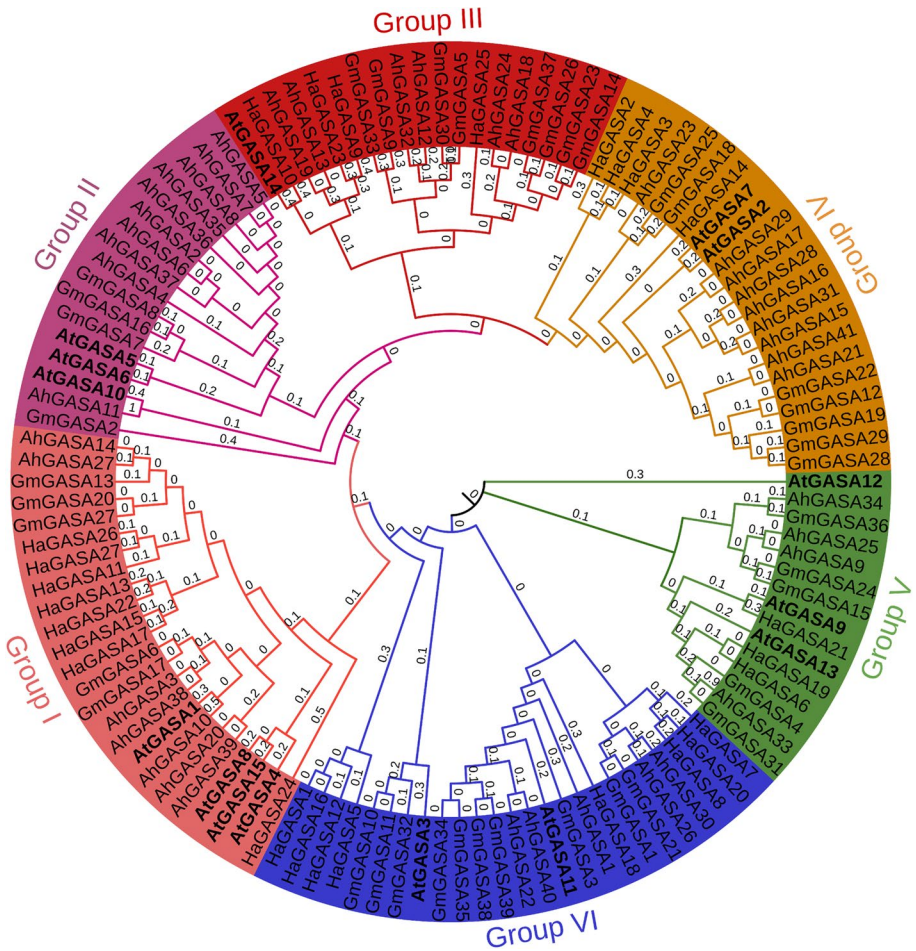

Fig. 2 Evolutionary tree of GASA genes from Arabidopsis, sunflower, peanut and soybean

TCC-box, CAT-box, HD-Zip, 3GCN4_motif, MSA-like, O2-site, and RY-element) have been assessed, with the AAGAA motif being the most dominant.

Prediction of targeted MiRNA sites

Since miRNAs play a crucial role in regulating gene expression, genome-wide research is focused on identifying potential miRNA targets [79]. Precise prediction and confirmation of miRNA targets can expose the molecular mechanisms behind different diseases and guide the creation of specific treatment methods [80]. The miRNAs targeting *HaGASA* genes showed lengths spanning 19 to 22 nucleotides. Twelve miRNAs were found that targeted *HaGASA10*, *HaGASA19*, *HaGASA21*, *HaGASA23*, and *HaGASA26* genes (Table 1S). Four miRNAs targeted *HaGASA10*, one miRNA targeted *HaGASA19*, one miRNA targeted *HaGASA21*, three miRNAs targeted *HaGASA23* and three miRNAs targeted *HaGASA26*.

Transcriptomic analysis

RNA seq analysis of HaGASA genes in response to drought

Ten genotypes of the sunflower were utilized in that experiment to observe genetic variation under irrigated and water deficit conditions. Based on RNA-seq analysis of ten genotypes, the *HaGASA2*, *HaGASA5*, *HaGASA6*, *HaGASA7*, *HaGASA10*, *HaGASA11*, *HaGASA12*, *HaGASA16*, *HaGASA17*, *HaGASA18*, *HaGASA19*, *HaGASA20*, *HaGASA21* and *HaGASA24* genes depicted significant variation in expression in response to drought (Fig. 10). The expression of *HaGASA2*, *HaGASA10*, and *HaGASA11* genes was significantly up-regulated, reflecting a prominent role in drought tolerance mechanisms. Conversely, significant down-regulation was identified in *HaGASA6*, *HaGASA17*, *HaGASA18*, *HaGASA21*, and *HaGASA24*, revealing their crucial significance as homeobox genes in maintaining optimal growth.

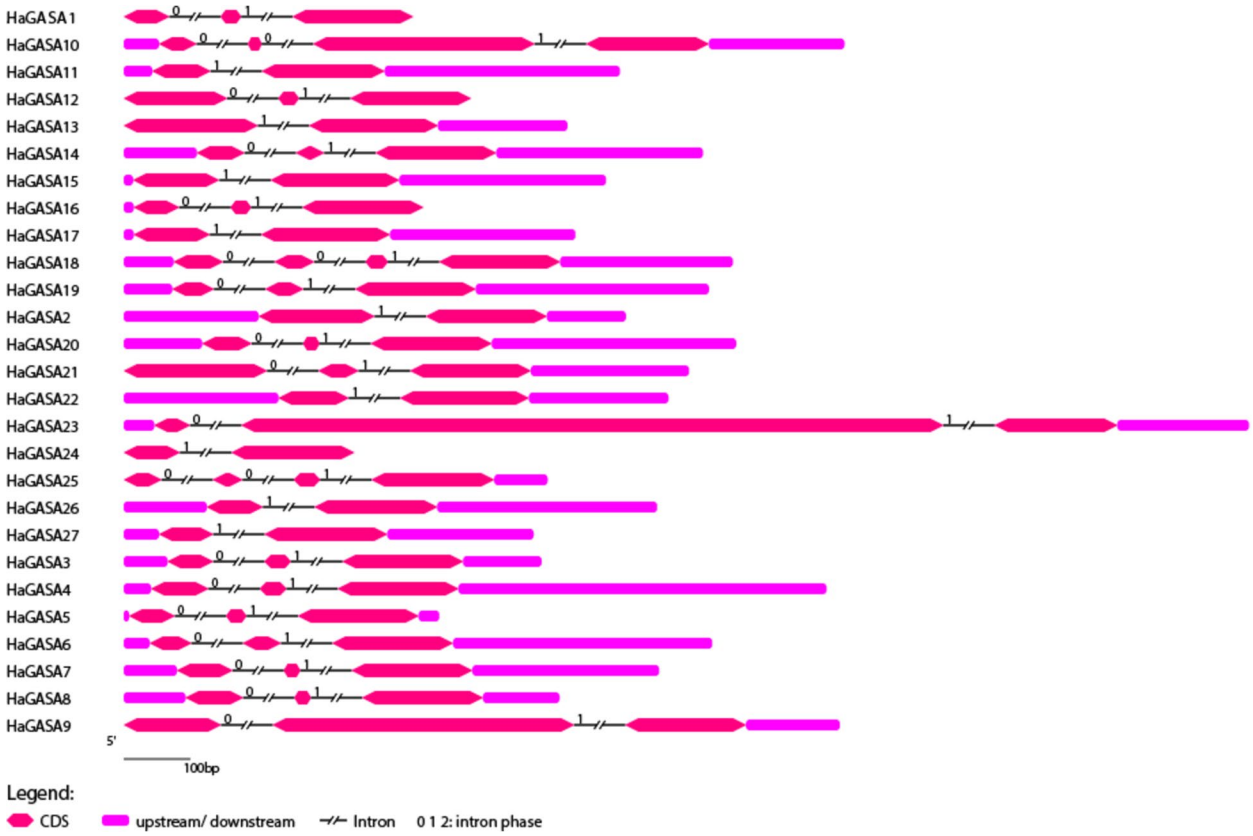

Fig. 3 Arrangement of coding and non-coding sequence of HaGASA genes

Transcriptomic analysis of HaGASA genes in various tissues

The *HaGASA* gene's transcriptomic expression was analyzed in distinct tissues. Nine genes (*HaGASA2*, *HaGASA8*, *HaGASA10*, *HaGASA11*, *HaGASA14*, *HaGASA17*, *HaGASA19*, *HaGASA21* and *HaGASA26*) out of 27 *HaGASA* genes depicted significant variation in expression in five distinct tissues (Fig. 11). The *HaGASA1*, *HaGASA12*, *HaGASA16*, and *HaGASA25* depicted zero expression in all different organs. Tissue-specific expression analysis revealed that *HaGASA2*, *HaGASA11*, *HaGASA17*, *HaGASA19*, *HaGASA21*, and *HaGASA26* exhibited peak expression levels in stem tissue. Conversely, *HaGASA8*, *HaGASA10*, and *HaGASA14* demonstrated maximal expression in flower, axil, and leaf tissues, respectively.

Discussion

In nature, plants are frequently exposed to multiple stressors, resulting in special and erratic circumstances [81]. Acute times of water scarcity have had detrimental effects on plant production and productivity in many parts of the world [82, 83]. Plants must adapt their metabolic and signaling responses to meet the unique physiological and developmental demands imposed by the array of these stresses. This involves modifications to photosynthesis, control over hormonal signaling pathways, molecular mechanism activation and antioxidant augmentation [84, 85]. A variety of stimuli, including drought stress, abscisic acid (ABA), reactive oxygen species (ROS), darkness, and high CO_2 concentrations in the surrounding air, can cause stomatal closure by inducing guard cells to release osmotic ions [86, 87]. The genes of the *GASA* family are crucial for the development of plants and reactions to the environment [88, 89]. Numerous functional investigations have shown that the *GASA* genes are critical in regulating plant growth, development, antibacterial activity, and pathogen defense mechanisms [39, 90]. Considering the lack of research on the involvement of these genes in sunflower stress tolerance, a comprehensive genome-wide identification of *GASA* genes has been undertaken to address the research gap.

The study systematically analyzed physiochemical features of the *HaGASA* protein family in the sunflower, like molecular weight, isoelectric point, grand average of hydropathy, and number of exons and intron (Table 1). The identified *GASA* genes exhibited relatively low molecular weight compared to other drought-related gene families such as *CCO* and *SAP* [45, 91]. All the genes

Fig. 4 Conserved motifs of HaGASA genes

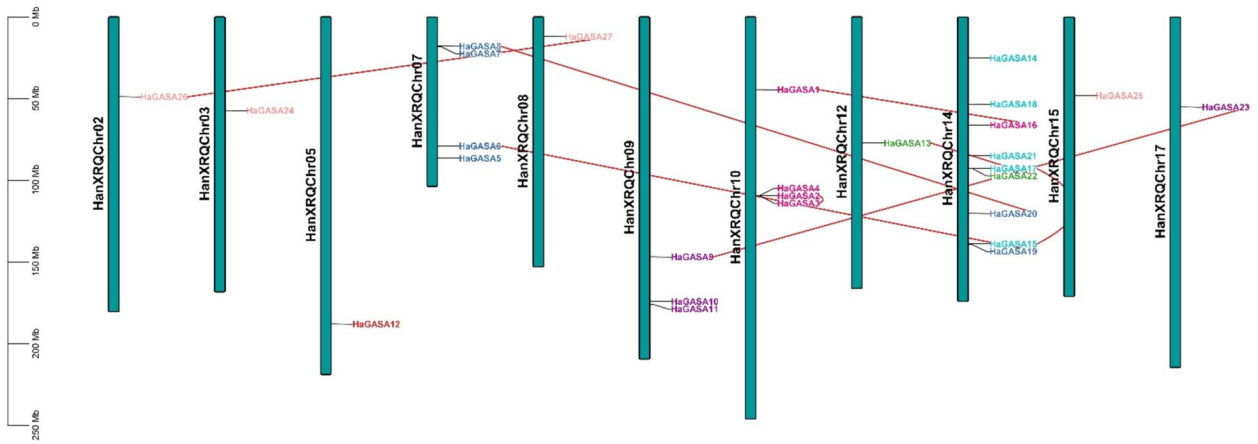

Fig. 5 Localization of HaGASA genes on the chromosomes of sunflower

were found to be unevenly present on specific chromosomes of the sunflower, like *Populus trichocarpa* and the potato, while *GASA* genes are consistently present on all chromosomes in Arabidopsis [12, 25, 29]. Most of the *HaGASA* proteins showed hydrophilicity, reflecting their negative GRAVY (Grand Average of Hydropathicity). This implies a strong attraction for water and the existence of net electrical charges over multiple pH levels [92]. All *HaGASA* genes were found to encode stable proteins, as each had an instability index below 40, which is the threshold for protein stability [93]. The location of *HaGASA* genes inside the cell was observed to predict cellular function based on the functioning of different organelles [94]. 88% of *HaGASA* genes were identified extracellularly, with 12% located in the nuclear and plasma membranes. Current in vivo research have

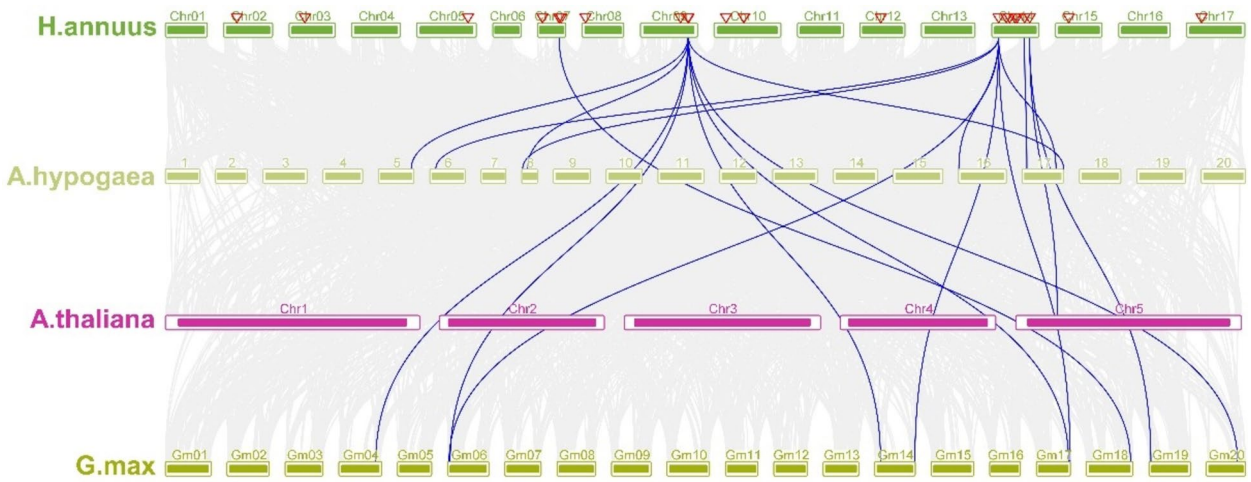

Fig. 6 Multiple synteny plot of sunflower, peanut, Arabidopsis and soybean depicting HaGASA collinear genes

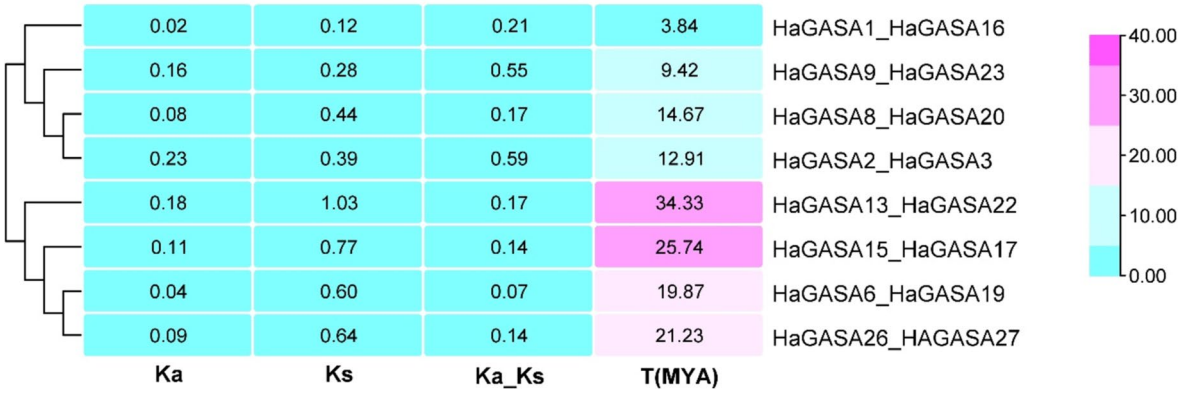

Fig. 7 Estimated times of gene duplication for various paralogous pairs of HaGASA genes derived from Ks and Ka values

revealed that *GASA* proteins are predominantly detected in extracellular spaces and cell walls [36, 95].

Comparing the *GASA* gene between various crops can be utilized to study the evolutionary connection of the *HaGASA* gene family [96]. The genes existing in similar clades can be speculated to perform the same function [97]. This can be utilized to predict the functioning of a less-studied gene through a highly-studied gene [98]. Therefore, functional genomics might be assisted by evaluating phylogeny connections [99]. A phylogenetic study was carried out, employing *GASA* gene sequences from *Arabidopsis thaliana* (*AtGASA*), *Arachis hypogaea* (*AhGASA*), *Glycine max* (*GmGASA*), and *H. annuus* (*HaGASA*).(Fig. 2). The *HaGASA* gene family was systematically divided into five different subgroups, demonstrating that genes located within the same clade are closely evolutionarily related. The *HaGASA* genes have functional resemblance with their *AtGASA*, *AhGASA*, and *GmGASA* counterparts in the same clades [20].

Previous research has revealed that the coding and non-coding sequences organization within a gene can be used to predict its evolutionary relationships with other genes [100, 101]. The majority of *HaGASA* genes comprised 2 to 4 conserved exons, while varied from 1 to 4 in *Citrus clementina* [26]. The *GASA* genes that shared similar exon-introns clustered within the same clade, indicating a close evolutionary link and a common ancestral origin (Fig. 3). The conserved motifs were evaluated which showed that motifs 1, 2 and 3 were common in all *HaGASA* genes while motif 9 was uniquely present in *HaGASA2* and *HaGASA21*(Fig. 4). This similarity increases the likelihood that these genes exhibit functional similarities depending on their conserved motif. The consistent existence of such motifs indicates they are crucial for controlling or operating genes [102]. This suggests these motifs primarily contribute to the *HaGASA* gene's behavior and regulation.

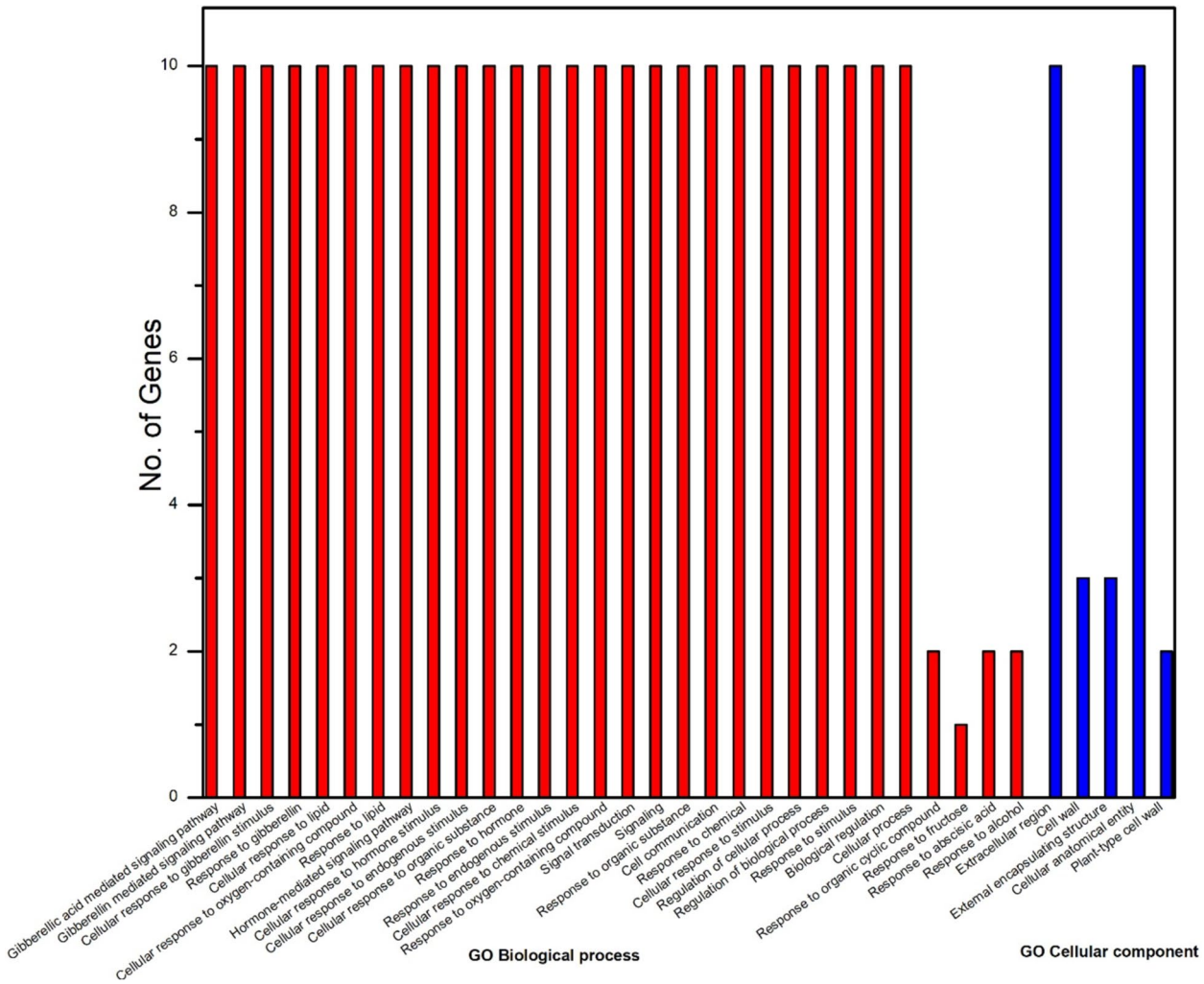

Fig. 8 Gene enrichment chart of HaGASA genes where GO biological process is indicated by red bars and GO cellular component is represented by blue bars

The evolution of a gene can be speculated via comparing the genomes of different species through comparative syntenic mapping [103]. A comparative genomic analysis was conducted among sunflower, Arabidopsis, peanut, and soybean to identify and characterize collinear blocks (Fig. 6). A high number of *HaGASA* genes copies were found in the genomes of the soybean and the peanut while Arabidopsis contained just a single copy of the *HaGASA* genes. This reveals that these crops might evolved through a common ancestor compared to arabidopsis. Gene expansion because of mutations and other environmental factors might cause the variation of *HaGASA* genes in the genomes of different species [104, 105].

The Ka/Ks ratio reveals crucial information on the selective pressures governing amino acid substitutions [106]. A Ka/Ks ratio below one signifies purifying selection that eliminates harmful mutations, whereas a ratio greater than one indicates positive selection which favors favorable mutations and drives adaptive evolution [107]. The Ka/Ks of *HaGASA* paralogous gene pairs ranged from 0.06 to 0.58, which is below one which indicates that during the process of evolution. All the gene pairs have undergone strong purification and positive selection took place at some sites (Fig. 7) [108].

A gene's transcriptional expression is highly dependent on the upstream promoter region of that gene [109, 110]. The promoter region of a protein contains elements that are specified for performing a specific function to handle various factors including plant growth and stress response [111, 112]. The analysis of *HaGASA* promoters depicted a vast array of elements that were responsive to light, growth, hormones and stress (Fig. 9). The *HaGASA* genes displayed a high density of cis-acting elements predominantly associated with stress

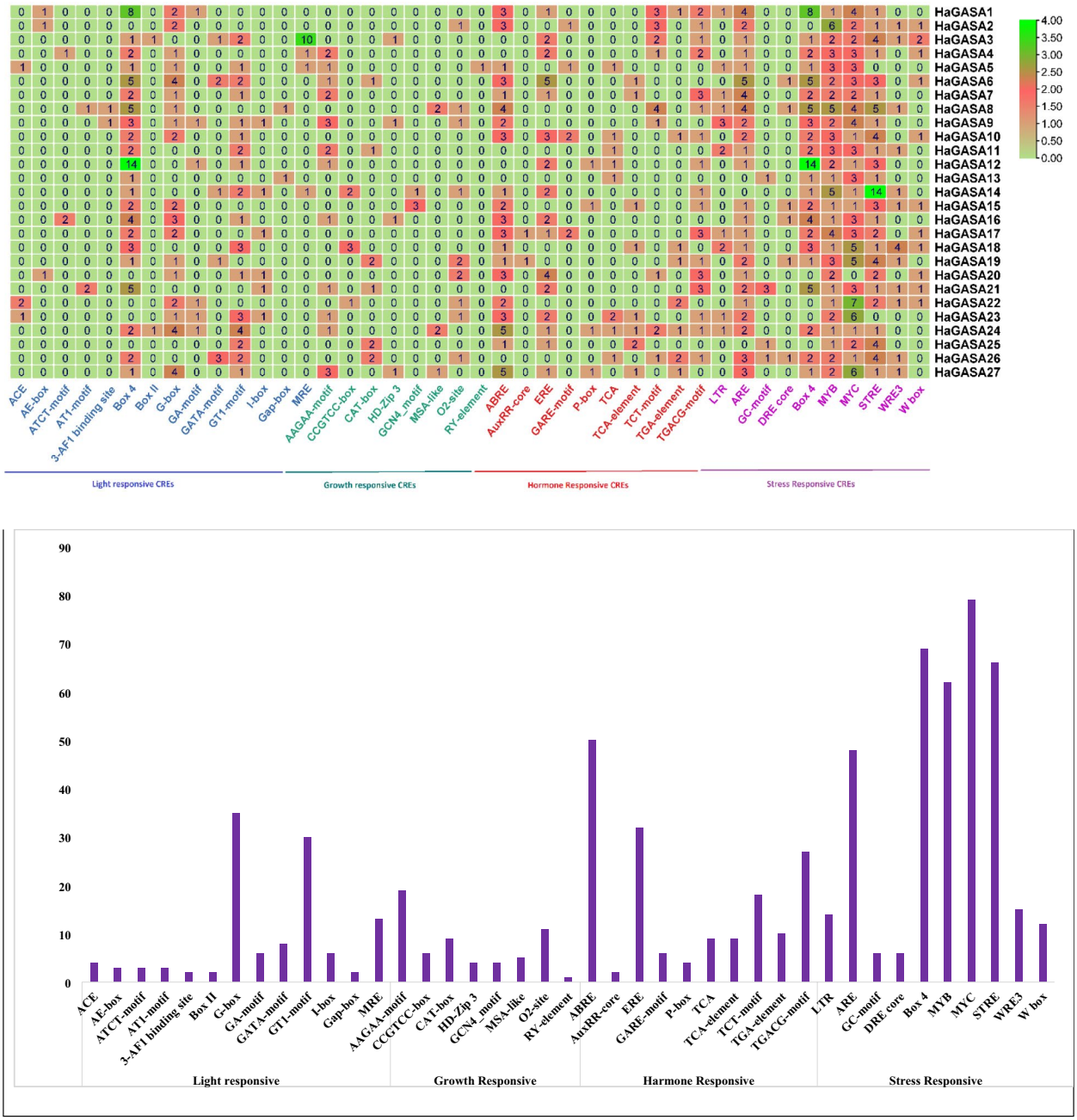

Fig. 9 Various CREs in the promoter region of HaGASA genes responsive to stress, hormones, light and growth

responsiveness, followed by those linked to hormonal signals, light, and growth-related processes. This pattern suggests that *HaGASA* genes may play a pivotal role in modulating responses to both biotic and abiotic stresses. These findings align with the observed behavior of *GASA* genes in *Citrus clementina* [26]. The *HaGASA* genes contained fewer growth-responsive elements and numerous hormone-responsive elements, may be more potentially involved in stress responses, and may participate in various hormonal signaling pathways.

A gene's transcriptional response is highly controlled by miRNAs targeting that gene [113]. The miRNAs are highly targeted in regulating particular biological functions [114]. Our research identified twelve microRNAs targeting five *HaGASA* genes. Such miRNAs function by inhibiting cleavage and translation processes, hence controlling gene expression Each miRNA targets an individual specific gene and multiple miRNAs may converge on a single target gene, potentially exerting a synergistic effect on its regulation [115–117]. Some

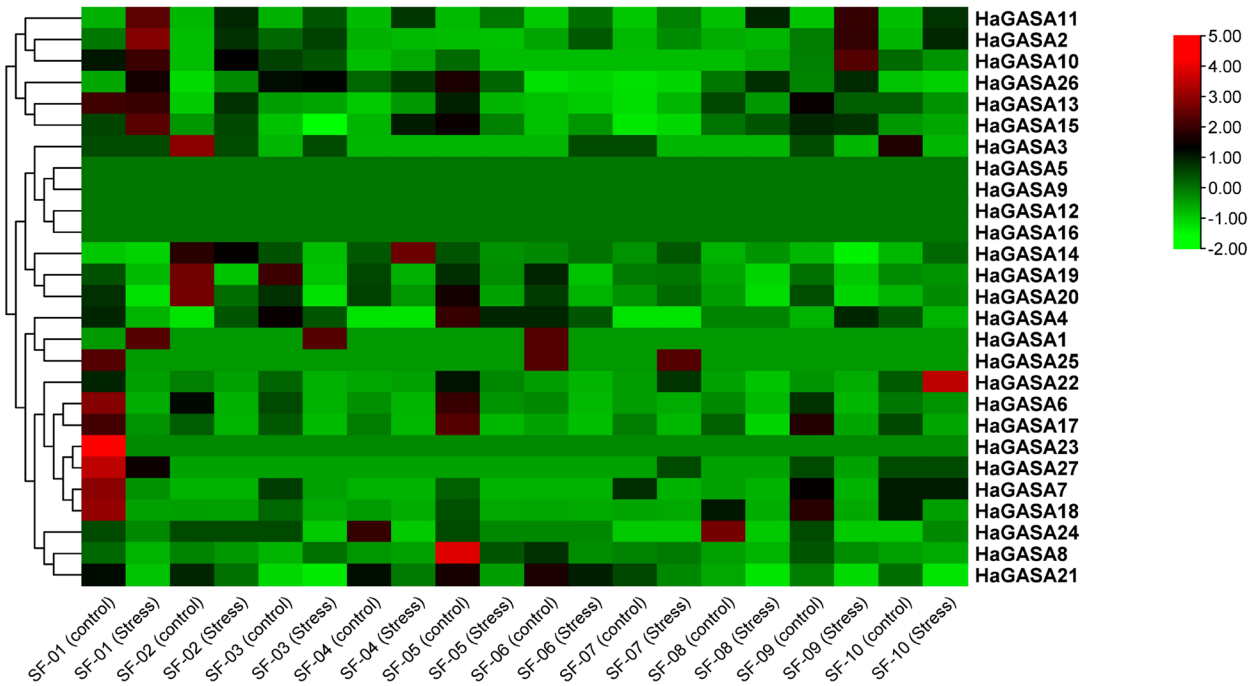

Fig. 10 Heatmap showing expression of HaGASA genes in ten genotypes of sunflower in response to drought

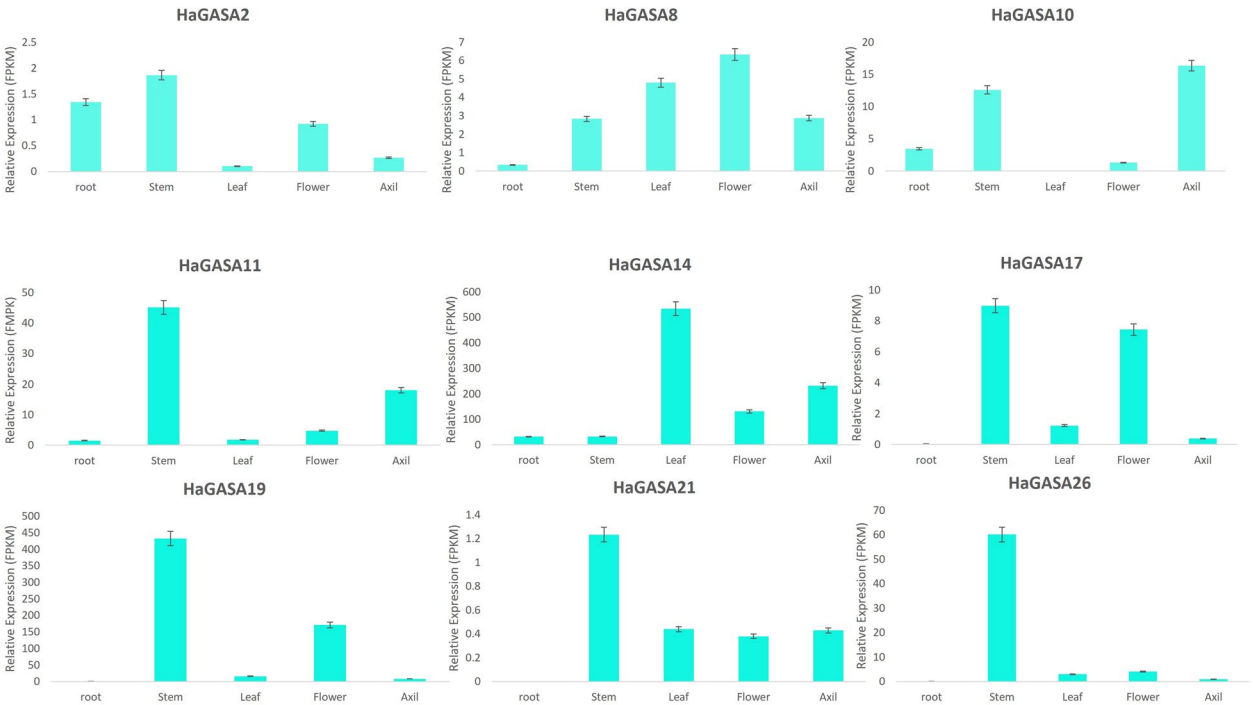

Fig. 11 Expression of nine HaGASA genes in five different organs of sunflower

microRNAs can inhibit the translation of their target genes, the primary mechanism of action for most miRNAs involves inducing the cleavage of messenger RNA. In our study, all discovered microRNAs downregulate gene expression by mRNA cleavage, predominantly lowering mRNA levels through degradation pathways, which leads to decreased protein output like the *CCO* gene in sunflower [91].

Among abiotic stresses, water scarcity is considered the most detrimental to plant development via inhibiting transpiration due to stomatal closure, ultimately leading to a significant reduction in yield [118, 119]. Transcriptomic data of ten sunflower genotypes (GSE145709) depicting the expression of *HaGASA* genes under water stress was utilized to identify potential *HaGASA* genes that show responsiveness to drought condition. The genes *HaGASA2*, *HaGASA10* and *HaGASA11* significant upregulation in response to drought, might be predicted to be involved in maintaining plant's various pathways through producing ABA [120] (Fig. 10). These genes can be can be further utilized in breeding projects to develop drought-resistant varieties of sunflower.

The biological function of a gene in plants can be predicted through its expression pattern in various organs [121, 122]. Transcriptomic analysis of *HaGASA* genes (GSE221055) was performed in five different organs to *HaGASA* genes activity in various plant organs corresponding to specific functions (Fig. 11). Certain *HaGASA* genes exhibit peak expression levels in the stem and leaf tissues, suggesting their critical involvement in regulating and sustaining various processes of plant's vegetative growth [123, 124]. The *HaGASA8* gene is most highly expressed in floral tissues, indicating a potential role in floral structure maintenance and development. The comprehensive investigation of the *HaGASA* gene family reveals their potential function in regulating both development processes and adaptive responses to drought.

Conclusion

Twenty-seven *GASA* genes were identified in the sunflower genome through in-silico analysis that were classified into five subgroups based on evolutionary analysis. The *HaGASA* genes were inconsistently present on the sunflower chromosomes. CREs responsive to stress, light, hormones and growth were identified in the promoters of *HaGASA* genes. Fourteen out of 27 *HaGASA* genes were found to be responsive to drought. *HaGASA2*, *HaGASA11*, *HaGASA17*, *HaGASA19*, *HaGASA21* and *HaGASA26* depicted high expression in the stem that might be involved in maintaining the growth of the sunflower. This research can be utilized as a foundation for further function research.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-024-10860-8.

Supplementary Material 1.

Acknowledgements
Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R459), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Authors' contributions
MUA helped to prepare the first draft of the manuscript. Bioinformatics analysis was handled by MAA, MLAH, KAA and MABZ. Every author helped to draft and edit the paper. MAU, KAA and MDFAK revised and MH supervised the manuscript. MLAH provided funding.

Funding
Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2024R459), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Availability of data and materials
The NCBI-GEO database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi using acc= GSE145709 and GSE221055) contains the datasets analyzed for this work.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
The authors say they have no conflicting agendas.

Competing interests
The authors have no competing interests.

Received: 6 July 2024 Accepted: 3 October 2024
Published online: 14 October 2024

References

1. Campos ML, et al. The role of antimicrobial peptides in plant immunity. J Exp Bot. 2018;69(21):4997–5011.
2. Sadelaji S, et al. Ib-AMP4 antimicrobial peptide as a treatment for skin and systematic infection of methicillin-resistant Staphylococcus aureus (MRSA). Iran J Basic Med Sci. 2022;25(2):232.
3. Liu L, et al. Small but powerful: RALF peptides in plant adaptive and developmental responses. *Plant Sci*. 2024;343:112085. https://doi.org/10.1016/j.plantsci.2024.112085.
4. Jangra R, et al. Duplicated antagonistic EPF peptides optimize grass stomatal initiation. Development. 2021;148(16):dev199780.
5. Caine RS, et al. An ancestral stomatal patterning module revealed in the non-vascular land plant Physcomitrella patens. Development. 2016;143(18):3306–14.
6. Mizuta Y, Higashiyama T. Chemical signaling for pollen tube guidance at a glance. J Cell Sci. 2018;131(2):jcs208447.
7. Shahin-Kaleybar B, et al. Isolation of cysteine-rich peptides from Citrullus colocynthis. Biomolecules. 2020;10(9):1326.
8. Silverstein KA, et al. Small cysteine-rich peptides resembling antimicrobial peptides have been under-predicted in plants. Plant J. 2007;51(2):262–80.
9. Ahmad B, et al. 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. Int J Mol Sci. 2020;21(3):1088.
10. Shi L, Olszewski NE. Gibberellin and abscisic acid regulate GAST1 expression at the level of transcription. Plant Mol Biol. 1998;38:1053–60.
11. Shaban M, et al. Genome-wide dissection, characterization, and expression profiling of cotton GASA genes reveal their importance in regulating abiotic stresses. 2021.
12. Zhang S, Wang X. Expression pattern of GASA, downstream genes of DELLA, in Arabidopsis. Chin Sci Bull. 2008;53(24):3839–46.

13. Ben-Nissan G, et al. GIP, a Petunia hybrida GA-induced cysteine-rich protein: a possible role in shoot elongation and transition to flowering. Plant J. 2004;37(2):229–38.
14. Aubert D, et al. Expression patterns of GASA genes in Arabidopsis thaliana: the GASA4 gene is up-regulated by gibberellins in meristematic regions. Plant Mol Biol. 1998;36:871–83.
15. Porto WF, Franco OL. Theoretical structural insights into the snakin/GASA family. Peptides. 2013;44:163–7.
16. Furukawa T, Sakaguchi N, Shimada H. Two OsGASR genes, rice GAST homologue genes that are abundant in proliferating tissues, show different expression patterns in developing panicles. Genes Genet Syst. 2006;81(3):171–80.
17. Fan S, et al. Comprehensive analysis of GASA family members in the Malus domestica genome: identification, characterization, and their expressions in response to apple flower induction. BMC Genomics. 2017;18:1–19.
18. Zhang K, et al. Genome-wide identification of GASA gene family in ten cucurbitaceae species and expression analysis in cucumber. Agronomy. 2022;12(8):1978.
19. Qiao K, et al. Identification, characterization, and expression profiles of the GASA genes in cotton. J Cotton Res. 2021;4:1–16.
20. Wu Y, et al. Comprehensive analysis of GASA family members in the peanut genome: identification, characterization, and their expressions in response to pod development. Agronomy. 2022;12(12):3067.
21. Cheng X, et al. Identification and analysis of the GASR gene family in common wheat (Triticum aestivum L.) and characterization of TaGASR34, a gene associated with seed dormancy and germination. Front Genet. 2019;10:980.
22. Ahmad MZ, et al. A genome-wide approach to the comprehensive analysis of GASA gene family in Glycine max. Plant Mol Biol. 2019;100:607–20.
23. Filiz E, Kurt F. Antimicrobial peptides Snakin/GASA gene family in sorghum (Sorghum bicolor): genome-wide identification and bioinformatics analyses. Gene Rep. 2020;20:100766.
24. Sun B, et al. Genome-wide identification and expression analysis of the GASA gene family in Chinese cabbage (Brassica rapa L. ssp. pekinensis). BMC Genomics. 2023;24(1):668.
25. Nahirñak V, et al. Genome-wide analysis of the Snakin/GASA gene family in Solanum Tuberosum Cv. Kennebec. Am J Potato Res. 2016;93:172–88.
26. Wu T, et al. Analysis of CcGASA family members in Citrus Clementina (Hort. Ex Tan.) By a genome-wide approach. BMC Plant Biol. 2021;21:1–18.
27. Yang M, et al. Genome-wide identification and characterization of Gibberellic Acid-stimulated Arabidopsis Gene Family in Pineapple (Ananas comosus). Int J Mol Sci. 2023;24(23):17063.
28. Li Z, et al. Genome-wide identification and characterization of GASA gene family in Nicotiana tabacum. Front Genet. 2022;12:768942.
29. Han S, et al. Genome-wide comprehensive analysis of the GASA gene family in Populus. Int J Mol Sci. 2021;22(22):12336.
30. Büyük I, et al. Identification and characterization of the Pvul-GASA gene family in thePhaseolus Vulgaris and expression patterns under salt stress. Turkish J Bot. 2021;45(7):655–70.
31. Su D, et al. Genome-wide characterization of the tomato GASA family identifies SlGASA1 as a repressor of fruit ripening. Hortic Res. 2023;10(1):uhac222.
32. Zimmermann R, Sakai H, Hochholdinger F. The gibberellic acid stimulated-like gene family in maize and its role in lateral root development. Plant Physiol. 2010;152(1):356–65.
33. de la Fuente JI, et al. The strawberry gene FaGAST affects plant growth through inhibition of cell elongation. J Exp Bot. 2006;57(10):2401–11.
34. Wang L, et al. OsGSR1 is involved in crosstalk between gibberellins and brassinosteroids in rice. Plant J. 2009;57(3):498–510.
35. Alonso-Ramírez A, et al. Evidence for a role of gibberellins in salicylic acid-modulated early plant responses to abiotic stress in Arabidopsis seeds. Plant Physiol. 2009;150(3):1335–44.
36. Li K-L, et al. GsGASA1 mediated root growth inhibition in response to chronic cold stress is marked by the accumulation of DELLAs. J Plant Physiol. 2011;168(18):2153–60.
37. Li X, et al. OsGASR9 positively regulates grain size and yield in rice (Oryza sativa). Plant Sci. 2019;286:17–27.
38. Wang H, et al. Transcriptome analyses from mutant Salvia miltiorrhiza reveals important roles for SmGASA4 during plant development. Int J Mol Sci. 2018;19(7):2088.
39. Segura A, et al. Snakin-1, a peptide from potato that is active against plant pathogens. Mol Plant Microbe Interact. 1999;12(1):16–23.
40. Mao Z, et al. The new CaSn gene belonging to the snakin family induces resistance against root-knot nematode infection in pepper. Phytoparasitica. 2011;39:151–64.
41. Khan S, et al. Sunflower oil: efficient oil source for human consumption. Emergent life Sci Res. 2015;1:1–3.
42. Fulda S, et al. Physiology and proteomics of drought stress acclimation in sunflower (Helianthus annuus L). Plant Biol. 2011;13(4):632–42.
43. Chen L, Yang J-y, Wang D. Phytoremediation of uranium and cadmium contaminated soils by sunflower (Helianthus annuus L.) enhanced with biodegradable chelating agents. J Clean Prod. 2020;263:121491.
44. Bashir SS, et al. Plant drought stress tolerance: understanding its physiological, biochemical and molecular mechanisms. Biotechnol Biotechnol Equip. 2021;35(1):1912–25.
45. Zhang C, et al. Genome-wide identification and evolution of the SAP gene family in sunflower (Helianthus annuus L.) and expression analysis under salt and drought stress. PeerJ. 2024;12:e17808.
46. Li W, et al. Genome-wide identification and comprehensive analysis of the NAC transcription factor family in sunflower during salt and drought stress. Sci Rep. 2021;11(1):19865.
47. Song H, et al. Genome-wide identification and expression analysis of the Dof gene family reveals their involvement in hormone response and abiotic stresses in sunflower (Helianthus annuus L). Gene. 2024;910:148336.
48. Li J, et al. Genome-wide identification of MYB genes and expression analysis under different biotic and abiotic stresses in Helianthus annuus L. Ind Crops Prod. 2020;143:111924.
49. Hussain M, et al. Genome-wide analysis of plant specific YABBY transcription factor gene family in carrot (Dacus carota) and its comparison with Arabidopsis. BMC Genomic Data. 2024;25(1):26.
50. Khatun K, et al. Genome-wide identification, genomic organization, and expression profiling of the CONSTANS-like (COL) gene family in petunia under multiple stresses. BMC Genomics. 2021;22:1–17.
51. Maqsood H, et al. Genome-wide identification, comprehensive characterization of transcription factors, cis-regulatory elements, protein homology, and protein interaction network of DREB gene family in Solanum lycopersicum. Front Plant Sci. 2022;13:1031679.
52. Horton P, et al. WoLF PSORT: protein localization predictor. Nucleic Acids Res. 2007;35(suppl2):W585-7.
53. Zhao L, et al. Genome-wide identification and analysis of the evolution and expression pattern of the HVA22 gene family in three wild species of tomatoes. PeerJ. 2023;11:e14844.
54. Ma Q, et al. Genomic analysis reveals phylogeny of Zygophyllales and mechanism for water retention of a succulent xerophyte. Plant Physiol. 2024;195.
55. Letunic I, Bork P. Interactive tree of life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 2024;52.
56. Hu B, et al. GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics. 2015;31(8):1296–7.
57. Dai Y, et al. Evolution and expression of the meprin and TRAF homology domain-containing Gene Family in Solanaceae. Int J Mol Sci. 2023;24(10):8782.
58. Li X, et al. Genome-wide identification of NAC transcription factor family in Juglans mandshurica and their expression analysis during the fruit development and ripening. Int J Mol Sci. 2021;22(22):12414.
59. Akter N, et al. Genome-wide identification and characterization of protein phosphatase 2 C (PP2C) gene family in sunflower (Helianthus annuus L.) and their expression profiles in response to multiple abiotic stresses. PLoS ONE. 2024;19(3):e0298543.
60. Lei Y, et al. Characterization and gene expression patterns analysis implies BSK family genes respond to salinity stress in cotton. Front Genet. 2023;14:1169104.
61. Song H, et al. Genome-wide characterization and comprehensive analysis of NAC transcription factor family in Nelumbo nucifera. Front Genet. 2022;13:901838.

62. Wang Y et al. Detection of colinear blocks and synteny and evolutionary analyses based on utilization of MCScanX. Nat Protoc. 2024:1–24.
63. Berendzen KW, et al. Cis-motifs upstream of the transcription and translation initiation sites are effectively revealed by their positional disequilibrium in eukaryote genomes using frequency distribution curves. BMC Bioinformatics. 2006;7:1–19.
64. Ho C-L, Geisler M. Genome-wide computational identification of biologically significant cis-regulatory elements and associated transcription factors from rice. Plants. 2019;8(11):441.
65. Luo X, et al. The evolution of the WUSCHEL-related homeobox gene family in dendrobium species and its role in sex organ development in D. chrysotoxum. Int J Mol Sci. 2024;25(10):5352.
66. Patel M, et al. Antioxidant effects and potential molecular mechanism of action of Diplocyclos Palmatus (L.) C. Jeffrey Fruits based on systematic network pharmacology with experimental validation. J Mol Struct. 2024;1313:138638.
67. Bu D, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic Acids Res. 2021;49(W1):W317-25.
68. Guo Z, et al. PmiREN: a comprehensive encyclopedia of plant miRNAs. Nucleic Acids Res. 2020;48(D1):D1114-21.
69. Tabassum N, et al. Genome-wide in-silico analysis of ethylene biosynthesis gene family in Musa acuminata L. and their response under nutrient stress. Sci Rep. 2024;14(1):558.
70. Gody L, et al. Transcriptomic data of leaves from eight sunflower lines and their sixteen hybrids under water deficit. OCL. 2020;27:48.
71. Wu Y, et al. Genome-wide analysis of TCP transcription factor family in sunflower and identification of HaTCP1 involved in the regulation of shoot branching. BMC Plant Biol. 2023;23(1):222.
72. Yan J, et al. Genome-wide association study and genetic mapping of BhWAX conferring mature fruit cuticular wax in wax gourd. BMC Plant Biol. 2022;22(1):539.
73. Sánchez D, et al. Exon-intron structure and evolution of the lipocalin gene family. Mol Biol Evol. 2003;20(5):775–83.
74. Cheng L, et al. Genome-wide identification, classification, and expression analysis of amino acid transporter gene family in Glycine max. Front Plant Sci. 2016;7:515.
75. Pond SLK, Poon AF, Frost SD. Estimating selection pressures on alignments of coding sequences. The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing. Cambridge, UK: Cambridge University Press; 2009. p. 419–90.
76. Newton IL, et al. Comparative genomics of two closely related Wolbachia with different reproductive effects on hosts. Genome Biol Evol. 2016;8(5):1526–42.
77. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nat Rev Genet. 2012;13(9):613–26.
78. Srinivasan C, et al. Addiction-associated genetic variants implicate brain cell type-and region-specific cis-regulatory elements in addiction neurobiology. J Neurosci. 2021;41(43):9008–30.
79. Huan T, et al. Genome-wide identification of microRNA expression quantitative trait loci. Nat Commun. 2015;6(1):6601.
80. Chen X, et al. MicroRNAs and complex diseases: from experimental results to computational models. Brief Bioinform. 2019;20(2):515–39.
81. Vos M, et al. The asymmetric response concept explains ecological consequences of multiple stressor exposure and release. Sci Total Environ. 2023;872:162196.
82. Mittler R. Abiotic stress, the field environment and stress combination. Trends Plant Sci. 2006;11(1):15–9.
83. Huang J, et al. Evaluation of regional estimates of winter wheat yield by assimilating three remotely sensed reflectance datasets into the coupled WOFOST–PROSAIL model. Eur J Agron. 2019;102:1–13.
84. Prasch CM, Sonnewald U. Simultaneous application of heat, drought, and virus to Arabidopsis plants reveals significant shifts in signaling networks. Plant Physiol. 2013;162(4):1849–66.
85. Zhou J, et al. H2O2 mediates the crosstalk of brassinosteroid and abscisic acid in tomato responses to heat and oxidative stresses. J Exp Bot. 2014;65(15):4371–83.
86. Song Y, Miao Y, Song CP. Behind the scenes: the roles of reactive oxygen species in guard cells. New Phytologist. 2014;201(4):1121–40. https://doi.org/10.1111/nph.12565.
87. Zhou Rong ZR, et al. High throughput sequencing of circRNAs in tomato leaves responding to multiple stresses of drought and heat. 2020.
88. Mu Y, et al. Cucumber CsBPCs regulate the expression of CsABI3 during seed germination. Front Plant Sci. 2017;8:459.
89. Yin L, et al. Wavelet analysis of dam injection and discharge in three gorges dam and reservoir with precipitation and river discharge. Water. 2022;14(4):567.
90. Su T, et al. Molecular and biological properties of snakins: the foremost cysteine-rich plant host defense peptides. J Fungi. 2020;6(4):220.
91. Sami A, et al. Genome-wide identification and in-silico expression analysis of CCO gene family in sunflower (Helianthus Annnus) against abiotic stress. Plant Mol Biol. 2024;114(2):34.
92. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157(1):105–32.
93. Gamage DG, et al. Applicability of instability index for in vitro protein stability prediction. Protein Pept Lett. 2019;26(5):339–47.
94. Itzhak DN, et al. Global, quantitative and dynamic mapping of protein subcellular localization. Elife. 2016;5:e16950.
95. Zhang S, et al. GASA5, a regulator of flowering time and stem growth in Arabidopsis thaliana. Plant Mol Biol. 2009;69:745–59.
96. Meyer RS, Purugganan MD. Evolution of crop species: genetics of domestication and diversification. Nat Rev Genet. 2013;14(12):840–52.
97. Gadagkar SR, Rosenberg MS, Kumar S. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J Experimental Zool Part B Mol Dev Evol. 2005;304(1):64–74.
98. Richardson R, et al. Meta-research: understudied genes are lost in a leaky pipeline between genome-wide assays and reporting of results. Elife. 2024;12:RP93429.
99. Barker D, Pagel M. Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. PLoS Comput Biol. 2005;1(1):e3.
100. Taft RJ, Pheasant M, Mattick JS. The relationship between non-protein-coding DNA and eukaryotic complexity. BioEssays. 2007;29(3):288–99.
101. Beer MA, Tavazoie S. Predicting gene expression from sequence. Cell. 2004;117(2):185–98.
102. Wray GA, et al. The evolution of transcriptional regulation in eukaryotes. Mol Biol Evol. 2003;20(9):1377–419.
103. Mayer KF, et al. Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell. 2011;23(4):1249–63.
104. Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. Annu Rev Ecol Evol Syst. 2009;40:481–501.
105. Yang X, et al. OsTTG1, a WD40 repeat gene, regulates anthocyanin biosynthesis in rice. Plant J. 2021;107(1):198–214.
106. Liu J, et al. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. Genome Biol. 2008;9:1–17.
107. Hurst LD. The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. 2002;18(9):486–7.
108. Massingham T, Goldman N. Detecting amino acid sites under positive selection and purifying selection. Genetics. 2005;169(3):1753–62.
109. Ayoubi TA, Van De Yen WJ. Regulation of gene expression by alternative promoters. FASEB J. 1996;10(4):453–60.
110. Cheng P, et al. Inclusion of root water absorption and reinforcement in upper bound limit stability analysis of vegetated slopes. Comput Geotech. 2024;169:106227.
111. Liu J-H, Peng T, Dai W. Critical cis-acting elements and interacting transcription factors: key players associated with abiotic stress responses in plants. Plant Mol Biology Report. 2014;32:303–17.
112. Yi J, et al. Assessing soil water balance to optimize irrigation schedules of flood-irrigated maize fields with different cultivation histories in the arid region. Agric Water Manage. 2022;265:107543.
113. Khraiwesh B, et al. Transcriptional control of gene expression by microRNAs. Cell. 2010;140(1):111–22.
114. Krützfeldt J, Poy MN, Stoffel M. Strategies to determine the biological function of microRNAs. Nat Genet. 2006;38(Suppl 6):S14-9.
115. Hausser J, Zavolan M. Identification and consequences of miRNA–target interactions—beyond repression of gene expression. Nat Rev Genet. 2014;15(9):599–612.
116. Yin L, et al. U-Net-LSTM: time series-enhanced lake boundary prediction model. Land. 2023;12(10):1859.

117. Zhao Y, et al. Characterizing uncertainty in process-based hydraulic modeling, exemplified in a semiarid Inner Mongolia steppe. Geoderma. 2023;440:116713.
118. Seleiman MF, et al. Drought stress impacts on plants and different approaches to alleviate its adverse effects. Plants. 2021;10(2):259.
119. Muhammad Asad U, Zia MAB. Morphological characterization of diverse wheat genotypes for yield and related traits under drought condition. Int J Nat Eng Sci. 2023;17(3):87–94.
120. Pan X, et al. Identification of ABF/AREB gene family in tomato (Solanum lycopersicum L.) and functional analysis of ABF/AREB in response to ABA and abiotic stresses. PeerJ. 2023;11:e15310.
121. Julca I, Tan QW, Mutwil M. Toward kingdom-wide analyses of gene expression. Trends Plant Sci. 2023;28(2):235–49.
122. Yin L, et al. Spatial and wavelet analysis of precipitation and river discharge during operation of the Three Gorges Dam, China. Ecol Ind. 2023;154:110837.
123. Nahirñak V, et al. Potato snakin-1 gene silencing affects cell division, primary metabolism, and cell wall composition. Plant Physiol. 2012;158(1):252–63.
124. Yin L, et al. U-Net-STN: a novel end-to-end lake boundary prediction model. Land. 2023;12(8):1602.

Publisher's note