Genome-Wide Identification and Characterization of SAC Domain-Containing Protein Family in Cotton

Xiaxuan Li
Zhengzhou University
Wei Chen
Institute of Cotton Research of CAAS
Shouhong Zhu
Institute of Cotton Research of CAAS
Yan Li
Institute of Cotton Research of CAAS
Jinbo Yao
Institute of Cotton Research of CAAS
Yongshan Zhang (13938698299@163.com)
Institute of Cotton Research of CAAS

Research article

Keywords: Phosphoinositides, SAC family, Bioinformatics, Synteny, Expression profile

Posted Date: October 6th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-80589/v1

Abstract

Background: Phosphoinositides(PIs) are important regulators of a diverse range of cellular functions. The Suppressor of Actin (SAC) domain-containing proteins are a class of phosphoinositide phosphatase involved in the synthesis of PIs. Though the cellular functions of SAC domain-containing proteins have been characterized in yeast, information of SAC genes in cotton is largely undefined.

Results: In the present study, 12,12and24 putative SAC genes were identified in the *G. ramondii, G. arboreum* and *G. hirsutum* respectively. Detailed gene information, including genomic organization, structural feature, conserved domain and phylogenetic relationship of the genes were systematically characterized. All SAC family members in cotton were divided into three clades, Group I, Group II and Group III, based on their sequence similarities and phylogenetic relationship. The SAC domains consist of seven highly conserved motifs that are believed to be important for the phosphoinositide phosphatase activities from yeast to animal. Expression analysis of *GhSAC* from Group II and Group III shared similar moderate pattern in different tissues and insensitive to different abiotic stresses. Different members in Group I showed different expressional profiles. Four genes (*GhSAC2.1A/GhSAC2.1D and GhSAC4.2A/GhSAC4.2D*) from Group I predominantly expressed in anther, pistil and petal. The results suggested the functional divergence among different groups and members of SAC in cotton.

Conclusions: Systematical analysis of the SAC gene family in cotton provided a solid foundation for further investigation of the biological functions of SAC genes.

Background

Phosphatidylinositol (PI) phosphates, which differ with regard to the presence or absence of phosphate groups on the available 3-, 4- and 5-hydroxy positions of the inositol head group [1], is a major and trace amounts of phospholipids in eukaryotic cells. They are collectively referred to as phosphoinositides (PIs), and exist as seven forms including PI(4)P, PI(3)P, PI(5)P, PI(4,5)P_2, PI(3,4)P_2, PI (3,5)P_2 and PI (3, 4, 5) P_3, one of which, PI(4,5)P_2 is known to be the precursor of the second messengers inositol 1, 4, 5-triphosphate and diacylglycerol, which are important in the activation of protein kinase C and the release of intracellular calcium [2]. Originally, PIs were thought to play a key role only in second messenger generation. However, a variety of researches about additional functions suggest that PIs are important regulators of a diverse range of cellular functions such as modulation of vesicle trafficking[3–5], cytoskeletal reorganization[6, 7], maintenance of vacuole morphology, activation of proteins[8], regulation of lipid storage, cell survival and cell proliferation[9–12].

PIs are synthesized by kinases and phosphatases which phosphorylate and dephosphorylate PI respectively. Based on the position of the phosphate that they hydrolyze, phosphoinositide phosphatases and inositol polyphosphate phosphatases are traditionally classified into four groups named 1-, 3-, 4-, or 5-phosphatase [8]. Among many phosphoinositide phosphatases, 5-phosphatase forms a fairly large family which is ulteriorly classified into four types according to their substrate specificity. Except for the type I 5-phosphatases that only use water-soluble inositol polyphosphate as substrates, the other three types are able to hydrolyzing phosphoinositides [8]. Recently, synaptojanin and inositide 5-phosphatases in which identified the SAC domain appeared to represent a novel group of phosphoinositide phosphatases. The SAC domain was originally found in the yeast phosphoinositide phosphatase Sac1p, which was identified in screens for “suppressor of actin” mutations [13] and suppressors of the defects caused by mutations of the Sec14 PI/phosphatidylcholine transfer protein [14]. Subsequently, the SAC domain was found in several proteins from yeast and animals. The SAC domain-containing proteins divided into two classes based on the specificity of the c-terminal amino acid sequences after the SAC domain [1]. The first class, which in addition to an N-terminal SAC domain, have all the domains associated with type II phosphoinositide 5-phosphatases, comprises mammalian synaptojanins and yeast Inp51p, Inp52p and Inp53p. The other class is represented by Sac1p and Fig. 4p in which the SAC domain is linked to a C-terminal region without any recognizable domains. This class includes yeast Fig. 4p and Sac1p which is the archetype of the Sac family of phosphatases and a quantity of uncharacterized proteins such as human (*Homo sapiens*) hSac1, hSac2 and hSac3. The C-terminal regions of the proteins in this class differ from one another and each has its own sequence specificity.

The association with phosphoinositide phosphatase activity with Sac domains had been identified. Through the detailed analysis of PI(4,5)P_2 hydrolysed by the 5-phosphatase synaptojanin, Chung suggested that the synaptojanin must exhibit the ability to dephosphorylate 4-phosphate groups[15]. Subsequently, characterization of mammalian synaptojanin and the yeast synaptojanin homologs Inp52p and Inp53p by Guo revealed a second phosphatase activity resides in the N-terminal SAC domain which was demonstrated to exhibit the activity capable of hydrolyzing phosphates from PI(3)P, PI(4)P and PI(3,5)P_2[16]. It is worth mentioning that the Sac phosphatases do not hydrolyse either PI(3,4)P_2 or PI(4,5)P_2, which contain adjacent phosphate groups. These appear that the SAC domain predominantly exhibits a lipid-specific phosphatidylinositol monophosphate phosphatase activity [1].

The Sac domain is approximately 400 amino acid residues in length and consists of seven highly conserved motifs that are believed to define the catalytic and regulatory regions of the phosphatase [1]. The sequence RXNCLDCLDRTN within the sixth motif is proposed to be the

catalytic core of the Sac domain phosphatases [1]. The CX_5R (T/S) motif which is thought to cradle the phosphate moiety is also present in a variety of metal-independent protein and inositide polyphosphate phosphatases [16]. Compare with Sac1p, Inp52p and Inp53p, Inp51p contains an incomplete CX_5R (T/S) motif does not exhibit phosphatase activity. Furthermore, the first conserved Asp residue is mutated into the RXNCXDCLDRTN sequence of the yeast *sac1- sac8* and *sac1- sac22* mutant alleles [17], which are thought to be the cause of the lack of phosphatase activity also indicated that the RXNCXDCLDRTN motif could well represent the catalytic core of the Sac phosphatases.

The cellular functions of Sac domain-containing proteins have been characterized, in particular Sac1p. Sac1p is an integral membrane protein [18] and plays an important role in ATP transport specifically in the endoplasmic reticulum [19] in which is Sac1p primarily localized [20–22]. Mutational analysis has demonstrated that Sac1p functions primarily to hydrolyze phosphate group from PI(4)P in vivo. Numerous researches about mutations of Sac1p indicated that Sac1p is involved in vesicle formation and transport [23–25], Golgi function, vacuole morphology [19] and actin cytoskeleton organization [1]. Figure 4p, the other yeast Sac domain-containing protein, in addition to showing 5-phosphatase activity [26] was required for the proper actin organization and cellular morphogenesis during the mating [27]. Inp5-phosphatases (Inp51p, Inp52p and Inp53p) overlap with each other while retaining some unique functions. Inp51p, as same as Inp52p, is clearly involved in endocytosis and regulation of the actin cytoskeleton under conditions of normal vegetative growth [1]. Except for Inp51p which exhibits only $PI(4,5)P_2$ 5-phosphatase activity, the others are in a position to be able to convert all of these PIs found in yeast into PI. As for Inp53p, Chang proposed that the protein may possess Golgi-to-vacuolar trafficking [28]. Moreover, several SAC domain-containing proteins from animals have been demonstrated to exhibit phosphoinositide phosphatase activities in vitro, but their cellular functions remain unknown [29–30].

Except $PI(3,4,5)P_3$, all phosphoinositides have been identified with plant cells. Some studies had suggested that phosphoinositides are involved in many important cellular activities such as osmotic regulation [31], plant defense response [32], vesicle trafficking [33, 34], pollen tube growth [35], and responses to stress and hormonal treatments [36–40]. However, much less is known about phosphoinositide phosphatases in plants. Although SAC phosphatases are essential regulators of PI-signaling network, little study has described regarding them and their possible biochemical and cellular functions in plants. In *Arabidopsis*, truncated *AtSAC1* has been proved to cause defects in cell morphogenesis and cell wall synthesis [41]. Gene expression analysis demonstrated that *AtSAC6* was predominantly expressed in the flowers and the expression was highly induced by salinity [42]. *AtSAC7* has been shown to be involved in root hairs growth [43]. Moreover, *AtSAC2-AtSAC5* have been characterized as an unknown subgroup of tonoplast-associated enzymes, was recently found to be involved in vacuolar morphology.

To characterize the molecular biology and evolution of the cotton SAC family and to understand its possible functions, it is necessary to identify its members and determine their expression patterns. In this report, we show that *G.hirsutum* genome contains 24 SAC domain-containing proteins, all of which belong to the class of Sac1p-like SAC proteins. We analyzed their gene structures, chromosomal locations, evolutionary relationships and expression patterns. Present analysis data shows that the GhSAC proteins fall into three subgroups based on their sequence homology and phylogenetic relationship. This is the first study to undertake a genome-wide analysis of *GhSACs*. These results provide valuable information on SAC genes in *G. hirsutum* and supply a framework to further studies to better understand the potential functions of SAC genes in cotton plants.

Results

Identification of SAC domain-containing proteins

HMMER searched was performed against the *T. cacao, V. vinifera, G. hirsutum, G. raimondii* and *G. arboreum* protein databases with SAC-domain PF02383 as a query. As a result, 6, 6, 29, 17, 13 putative SAC genes were identified initially. Meanwhile, all Arabidopsis SAC protein sequences were used as queries for TBLASTN. We checked all the sequences by Interpro online tool to search the SAC domain. Ultimately,6□6□24□12□12□SAC domain-contained proteins were identified from the above five genomes respectively. All *SAC* genes in *G. hirsutum* are designated as *GhSAC* and named according to the order of the closest orthologues in *Arabidopsis* [44]. The accession number, chromosome distribution, protein molecular weight and length of the *GhSAC* genes were listed in Table 1. By comparison of number of genes in the three closely related species, *SAC* gene family members in *G. hirsutum* showed an obvious expansion of number of genes.

Table 1
Nomenclature of SAC genes

Gene Name	Chromosomes	Start	End	Gene Length(bp)	Gene ID	Protein(aa)	CDS(bp)	Locus
GhSAC1.1A	A02	34096061	34112341	16280	GH_A02G1007.1	908	2727	+
GhSAC1.1D	D02	24480105	24496398	16293	GH_D02G1055.1	908	2727	+
GhSAC2.1A	A05	12818773	12826690	7917	GH_A05G1394.1	812	2439	+
GhSAC2.2A	A06	125121473	125129160	7687	GH_A06G2227.1	799	2400	-
GhSAC2.3A	A10	5084206	5091417	7211	GH_A10G0522.1	807	2424	+
GhSAC2.1D	D05	11741598	11749565	7967	GH_D05G1409.1	812	2439	+
GhSAC2.2D	D06	64016931	64024566	7635	GH_D06G2261.1	799	2400	-
GhSAC2.3D	D10	4755877	4763002	7125	GH_D10G0550.1	807	2424	+
GhSAC3.1A	A06	22413672	22420686	7014	GH_A06G0875.1	827	2484	-
GhSAC3.1D	D06	15193386	15200340	6954	GH_D06G0859.1	827	2484	-
GhSAC4.1A	A07	1784997	1791282	6285	GH_A07G0178.1	834	2505	-
GhSAC4.2A	A13	105624342	105631137	6795	GH_A13G2182.1	828	2487	+
GhSAC4.1D	D07	1777332	1783572	6240	GH_D07G0189.1	834	2505	-
GhSAC4.2D	D13	59684021	59690829	6808	GH_D13G2164.1	828	2487	+
GhSAC6.1A	A10	19377649	19381844	4195	GH_A10G0992.1	444	1335	+
GhSAC6.1D	D10	11132719	11138798	6079	GH_D10G0964.1	599	1800	-
GhSAC7.1A	A02	108027042	108033476	6434	GH_A02G2039.1	596	1791	+
GhSAC7.1D	D03	171041	177362	6321	GH_D03G0023.1	596	1791	-
GhSAC8.1A	A04	76778183	76783215	5032	GH_A04G1114.1	602	1809	-
GhSAC8.1D	D04	47908493	47913530	5037	GH_D04G1457.1	628	1887	-
GhSAC9.1A	A02	642432	657397	14965	GH_A02G0080.1	1930	5793	+
GhSAC9.2A	A09	79901119	79918983	17864	GH_A09G2310.1	1630	4893	+
GhSAC9.1D	D02	688844	703749	14905	GH_D02G0086.1	1927	5784	+
GhSAC9.2D	D09	49049613	49062591	12978	GH_D09G2248.1	1630	4893	+
GaSAC1	chr03	39716438	39732638	16200	Ga03G1088.1	908	2727	+
GaSAC2.1	chr05	12981653	12989776	8123	Ga05G1465.1	812	2439	+
GaSAC2.2	chr06	130666881	130674547	7666	Ga06G2488.1	799	2400	+
GaSAC2.3	chr10	124288924	124296135	7211	Ga10G2534.1	809	2430	-
GaSAC3	chr06	20549396	20556110	6714	Ga06G0884.1	827	2484	+
GaSAC4.1	chr07	2003435	2009716	6281	Ga07G0186.1	834	2505	-
GaSAC4.2	chr13	118864840	118871627	6787	Ga13G2361.1	828	2487	+
GaSAC6	chr10	108262526	108268588	6062	Ga10G1985.1	599	1800	-
GaSAC7	chr02	305956	312394	6438	Ga02G0025.1	596	1791	-
GaSAC8	chr04	11351705	11356764	5059	Ga04G0609.1	622	1869	+
GaSAC9.1	chr03	639823	657965	18142	Ga03G0085.1	1939	5820	+
GaSAC9.2	chr09	81728068	81745949	17881	Ga09G2424.1	1630	4893	+

Gene Name	Chromosomes	Start	End	Gene Length(bp)	Gene ID	Protein(aa)	CDS(bp)	Locus
GrSAC1	chr05	22714765	22731541	16776	Gorai.005G115800.1	908	2727	+
GrSAC2.1	chr09	10929781	10938648	8867	Gorai.009G144100.1	883	2652	+
GrSAC2.2	chr10	60633465	60641801	8336	Gorai.010G235900.1	799	2400	-
GrSAC2.3	chr11	4476638	4484614	7976	Gorai.011G056600.1	811	2436	+
GrSAC3	chr10	14796621	14804827	8206	Gorai.010G092400.1	827	2484	-
GrSAC4.1	chr01	1641103	1648424	7321	Gorai.001G017800.1	834	2505	-
GrSAC4.2	chr13	54202495	54209306	6811	Gorai.013G222100.1	828	2487	+
GrSAC6	chr11	10841283	10848063	6780	Gorai.011G097800.1	599	1800	-
GrSAC7	chr03	159037	166166	7129	Gorai.003G002700.1	596	1791	-
GrSAC8	chr12	26630443	26635722	5279	Gorai.012G115300.1	605	1818	-
GrSAC9.1	chr05	705299	715171	9872	Gorai.005G010100.1	1611	4836	+
GrSAC9.2	chr06	48157290	48171812	14522	Gorai.006G232600.1	1630	4893	+
VviSAC1	chr14	8378602	8409165	30563	VIT_214s0081g00460.1	614	1845	-
VviSAC3	chr09	414338	425305	10967	VIT_209s0002g00590.1	850	2553	+
VviSAC4	chr11	427858	442383	14525	VIT_211s0016g00440.1	835	2508	+
VviSAC7	chr04	20669057	20711649	42592	VIT_204s0044g00030.1	599	1800	+
VviSAC8	chr08	7811659	7821276	9617	VIT_208s0105g00480.1	608	1827	-
VviSAC9	chr05	24160942	24192208	31266	VIT_205s0094g00850.1	1644	4935	+
TcSAC1	Chr08	19053519	19066755	13236	Thecc.08G188300.1	913	2739	+
TcSAC2	Chr06	21734116	21743385	9269	Thecc.06G127600.1	814	2442	-
TcSAC4	Chr09	3580273	3588581	8308	Thecc.09G070200.1	844	2532	+
TcSAC6	Chr01	304575	312871	8296	Thecc.01G006400.1	598	1794	-
TcSAC8	Chr05	667593	673314	5721	Thecc.05G013700.1	590	1770	+
TcSAC9	Chr04	26260218	26276209	15991	Thecc.04G163000.1	1672	5016	-

Table 2
The cis-element analysis of *GhSACs* promoters

Gene	A	B	C	D	E	F	G	H	I	J
GhSAC1.1A	5	1	1	0	2	2	1	1	0	0
GhSAC1.1D	4	3	1	0	1	2	1	1	1	0
GhSAC2.1A	7	4	2	1	0	4	3	3	2	0
GhSAC2.2A	4	0	0	0	0	0	5	5	0	0
GhSAC2.3A	2	0	0	0	0	0	1	1	0	2
GhSAC2.1D	6	4	0	1	0	5	3	3	1	2
GhSAC2.2D	6	6	2	1	0	4	4	4	0	0
GhSAC2.3D	4	1	0	1	1	1	2	2	0	2
GhSAC3.1A	2	1	2	2	0	0	0	0	2	1
GhSAC3.1D	3	4	5	2	1	2	1	1	2	0
GhSAC4.1A	4	1	5	2	0	2	1	1	0	2
GhSAC4.2A	1	1	0	0	0	1	0	0	1	1
GhSAC4.1D	6	4	5	1	0	3	1	1	0	1
GhSAC4.2D	1	6	0	0	0	4	0	0	1	1
GhSAC6.1A	0	0	0	0	1	0	1	1	1	0
GhSAC6.1D	2	1	0	1	1	1	1	1	0	0
GhSAC7.1A	2	1	2	0	0	2	0	0	3	0
GhSAC7.1D	2	2	2	0	0	3	0	0	2	0
GhSAC8.1A	3	0	3	2	1	0	0	0	1	0
GhSAC8.1D	4	0	3	3	2	0	0	0	1	0
GhSAC9.1A	4	0	2	1	2	0	4	4	0	0
GhSAC9.2A	3	3	4	0	0	2	1	1	0	0
GhSAC9.1D	7	4	1	1	1	3	1	1	0	0
GhSAC9.2D	3	2	2	1	0	2	2	2	0	0

Phylogenetic analysis of the GhSACs

We constructed a phylogenetic tree from a multiple alignment of SAC protein sequences, comprising 6 TcSACs from *T. cacao*, 10 VviSACs from *V. vinifera*, 12 GaSACs from *G. arboreum*, and 9 AtSACs from *Arabidopsis*. The phylogenetic analysis revealed evolutionary origin for these genes as well as more recent duplications. The SAC proteins were clustered into three groups (Fig. 1), as previously suggested [42]. Genes from these species are found in all three groups, suggesting that the higher plant species have at least one gene in each of the three groups.

Our phylogenetic reconstruction showed that the SAC family in cotton diversified after the common ancestor of cotton and *Arabidopsis* because SAC genes of group I and group II in *G. arboretum* were obviously more than in *Arabidopsis*. And most of the SAC proteins from the diploids had orthologs in the allotetraploid *G. hirsutum*, which derived from a hybridization of A group and D group genome ancestors (Additional file 1). The short branches separating the paralogs suggested that the hybridization event occurred relatively recently [45].

Chromosome Localization And Synteny Analysis Of Sac Genes

To determine chromosome distribution and gene duplication of the *SAC* genes,all the *SAC* genes in *G. hirsutum* were mapped to approximate chromosome positions (Fig. 2). These twenty-five *GhSAC* genes were distributed among the 17 chromosomes unevenly. Except for A1, A3, A8,

A11, A12, D1, D8, D11 and D12, all chromosomes harbor at least one of the SAC genes. 12 and 12 *SAC* genes were found to located at the A-subgenome and D-subgenome respectively.

To further infer the phylogenetic mechanisms of SAC family, we constructed syntenic maps of *T. cacao* with *G. raimondii* and *V. vinifera* (Fig. 3). A total of 7 *GrSACs* and 4 *TcSACs* genes showed syntenic relationship with those in *T. cacao* and *V. vinifera*, respectively. *TcSAC2* and *TcSAC4* were found to be associated with more than one syntenic gene pairs between *G. raimondii* and *T.cacao* SAC genes, guessed that these genes may have played an important role of *SAC* gene family during evolution. In addition, *VviSAC9*/ *TcSAC9* gene pair identified between *T.cacao* and *V. vinifera* were not found between *G. raimondii* and *T. cacao*, which may indicate that this orthologous pair lost after the divergence of *G. raimondii* and *T.cacao* from their ancestors.

Gene structures and conserved domain of GhSACs

Gene structure analysis is important for studying genetic evolution. First, we mapped the domain structure by IBS software(version v1.0) (Fig. 4). Then, to understand the evolutionary relationship of SAC protein in *G. hirsutum*, we constructed the unrooted tree based on the alignments of full-length SAC protein sequences using MJ method of MEGA X. The 25 SAC proteins in *G. hirsuhum* were divided into three distinct groups (from I to III). Group I consist of the maximum number 14 of *GhSACs*, while group III contains only four *GhSACs*. The genomic sequence of the *GhSACs* genes ranged from 4195 bp to about 17 kb. To obtain further gene structure information, we compared the coding sequence with the genomic sequence of all *GhSAC* genes (Fig. 5). Different introns (from 6 to 19) were observed among the *GhSAC* genes. The genes possess maximum number of introns were in group II. The GhSAC proteins gene clusters that were divided into the same group exhibited similar structure. We used MEME to detect conserved motif in the *GhSAC* family. There were some differences between the groups. 20 conserved motifs were scattered among each *GhSAC* family (Fig. 5). All of the GhSAC proteins shared the same three motifs: M1, M2 and M3 these motifs together compose the SAC domain which was characteristic for all *GhSAC* family members.

The SAC domains of SAC proteins yeast and animal proteins are approximately 400 amino acids in length and consists of seven highly conserved motifs which appear to important for the phosphatase activities [1]. To examine in detail the motif organization of the SAC domains of the GhSAC proteins, we compared the SAC domain sequences between Sac1p and the GhSAC proteins and created the seven conserved motifs by the Weblogo online tools(Fig. 6A). Meanwhile, characteristic transmembrane motifs which followed by SAC domains in GhSAC proteins of Group II except GhSAC6.1A were also created.(Fig. 6B)

Sequence analysis showed that the GhSAC proteins except Group III contain all seven conserved motifs found in Sac1p (Additional file 3). The sixth conserved region contains a highly conserved $CX_5R(T/S)$ motif, which was identified as the catalytic motif in many metal-independent proteins and inositide polyphosphate phosphatases in previous reports. However, the putative catalytic core sequence RXNCXDCLDRTN located in motif VI is completely conserved among the GhSAC proteins (except these in Group III). This result suggests that GhSAC proteins may have SAC domain functions similar to those of yeast and animals.

In addition, we found that SAC proteins in subgroup III seemed to lack motif VII. However, in their place is a putative WW domain. WW domains have been shown to be involved in protein-protein interactions by recognizing Pro-containing ligands[46], and they are considered to be the smallest protein domain involved in protein-protein interactions. The WW domain is a short conserved region in a number of unrelated proteins, which folds as a stable, triple stranded beta-sheet. This short domain of approximately 40 amino acids, may be repeated up to four times in some proteins [47–49]. The name WW or WWP derives from the presence of two signature tryptophan residues that are spaced 20–23 amino acids apart and are present in most WW domains known to date, as well as that of a conserved Pro. It is frequently associated with other domains typical of proteins in signal transduction processes. The putative WW domain of these GhSAC proteins in Group III contained all the features typical of identified WW domains, such as the two Trp residues separated by 22 residues, and the presence of other conserved residues including the essential aromatic doublet and Pro. None of the other GhSAC proteins contains a putative WW domain. The functional significance of the putative WW domain in GhSACs of Group III remains to be investigated.

Cis-element analysis in the promoter regions of GhSAC **genes**

To identify the putative cis-acting regulatory elements, 2000 bp of sequence upstream from the start codon was isolated. Ultimately, we identified 44 different regulatory elements which divided into two main types: light responsive elements and hormone responsive elements from the promoter regions of *GhSACs.* (Table. 2)

Light responsive elements, including Box 4, G-Box, GT1-motif, GATA-motif and MRE, were enriched in the upstream promoter regions of *GhSAC* genes. Box 4, part of a conserved DNA module involved in light responsiveness, was the most abundant light responsive element in the promoters of *GhSAC* genes. The genes, except *GhSAC6.1A*, contained at least one Box 4 element. In addition, 19 members contained a G-Box element, 17 members contained a GT1-motif element, whereas 15 members contained a GATA-motif element. Then, we hypothesized

that light could induce the expression of *GhSAC* genes through their responsive cis-acting elements, further regulating the balance between reproductive and vegetative growth.

The other important type of cis-acting elements in the upstream regions of *GhSAC* genes are plant hormone-responsive elements. In total, nine types of elements were found that respond to five respective kinds of plant hormones. These regulatory elements included ABA-responsive elements (ABREs), MeJA-responsive elements(TGACG-motifs and CGTCA-motifs), salicylic acid responsive elements(TCA-elements), auxin-responsive element(TGA-elements). This indicates that *GhSAC* genes may respond to ABA, SA and JA.

Expression profile of GhSACs

To understand expression patterns of these 25 *GhSAC* genes in *G. hirsutum*, we used publicly available transcriptome data to assess the expression of different tissues and organs. The analysis (Fig. 7) revealed that four *GhSAC* genes (*GhSAC2.1A/GhSAC2.1D/GhSAC4.2A/GhSAC4.2D*) predominantly expressed in flowers, whereas the expression of other genes was not significantly altered in different tissues and two genes (*GhSAC2.2A/GhSAC2.2D*) were not expressed in all tissues and organs. In addition, the expression of *GhSAC* genes were not significantly altered under different abiotic stresses conditions, i.e. cold, heat, salt and drought (Addition file 6). We also performed RT-PCR to confirm the expression levels of four *GhSACs* in different tissues, including roots, stems, leaves, bracts, sepals, receptacles, petals, pistils, anthers. There was very high sequence similarity within these *GhSACs* CDSs of A-subgenome and D-subgenome, so primers were designed to detect the transcription levels of genes both in A- and D-subgenome. As shown in Fig. 8, *GhSAC2.1* and *GhSAC4.2* genes were predominantly expressed in stigmas and stamens with little expression in other organs while *GhSAC7.1* and *GhSAC9.2* were expressed in all organs examined. All these genes had a relatively lower level of expression in roots, stems and leaves. These results suggest that *GhSAC* genes have diverse expression patterns and some genes may play dominant roles in particular organs.

Discussion

With the increasing research in genomes, comparative genomics methods are used to study gene families, which is one of the hot research topics for several species. The SAC domain-containing protein gene was first identified in the yeast (*Saccharomyces cerevisiae*) named Sac1p phosphoinositide phosphatase protein. Although several other SAC domain-containing proteins from animals possess phosphoinositide phosphatase activities in vitro, their cellular functions remain unknown, in addition, much less is researched about these in plants [29, 30]. 9 *SAC* genes are identified in Arabidopsis [42], five in yeast [1] and five in human beings [30], however, the *G. hirsutum* genome has 24 members of the *SAC* gene family which is obviously much more than above. Zhong reported a genome-wide analysis of the SAC gene family members in *Arabidopsis* [42]. They discussed the number, classification, structure of genes and presented a basic analysis of the conserved motifs in SAC proteins.

In this study, we identified 24 *SAC* genes in the *G. hirsutum* genome, where 12 genes belong to the A subgenome and 12 genes to the D subgenome. Compared with other plants *SACs* (9 *SAC* family genes have been identified in *Arabidopsis*, 6 in *T. cacao*, 6 in *V. vinifera*, 12 in A group and 12 in D group), the *GhSAC* family is the largest with 24 phylogenetically expanded genes. The striking expansion and diversification of the *GhSAC* family genes probably suggests that these *SACs* play crucial roles in the physiological maintenance in *G. hirsutum*, which are same as Sac1p in yeast. We also noticed that *SAC* genes in AD genome were equaled to the sum of these in A genome and D genome. This result may be associated with the gene duplications in the evolution of AD genome from their diploid ancestors.

Although genes within a family evolve from multiple mechanisms, a comprehensive phylogenetic and structural analysis can offer insight into the evolutionary origins of, and relationships among, different isoforms [50]. Based on previous sequence similarities and phylogenetic relationship analyses [42], the AtSAC proteins have been divided into three subgroups. Our phylogenetic analysis of *Arabidopsis, T. cacao, V. vinifera* and cotton genes corroborated this classification and inferred that higher plant species have at least one gene in each of the three groups.

The existing research findings have demonstrated that the SAC domains of several proteins from yeast and human exhibit different specificities toward different phosphoinositides. For example, Sac1p, which contain the SAC domains, exhibit a broader-specificity phosphatase activity capable of hydrolysing phosphate from PI(3)P, PI(4)P, and PI(3,5)P_2 [16][29][51], whereas hSac2 possessed a 5-phosphatase activity toward PI(4,5)P_2 and PI(3,4,5) P_3 [30]. In plant cells, six forms of phosphoinositides have been detected [42]. Because GhSACs except these in subgroup III contain all seven conserved motifs, which believed to be important for the phosphatase activities of yeast and animal SAC proteins, we can speculate that GhSACs may function as phosphoinositide phosphatases. Moreover, the facts that the *G. hirsutum* genome contains 24 *SAC* genes belonging to three subgroups and suggest that different GhSACs might possess different substrate specificities, and, therefore, they may regulate the metabolism of different phosphoinositides in the phosphoinositide pool, which in

turn influences diverse cellular processes [42]. Definite proof of such an activity awaits the biochemical and functional characterization of the GhSAC proteins.

Gene expression pattern can provide important clues about gene functions, which are believed to be associated with divergence in the promoter region [52]. Cis-acting regulatory elements contained in gene' promoter regions play key roles in conferring the developmental regulation of gene expression. A total of 20 different types of light responsive element were identified in the promoter regions of *GhSAC* gene family via cis-element analysis. We found that light responsive elements were abundant in the promoters of *GhSACs* in each group, while the number of light responsive elements in the promoter regions of *GhSACs* in group I varies greatly, with the maximum of 19 and the minimum of only 2, which did not occur in the other groups. Therefore, we speculated that the *GhSAC* family genes were generally sensitive to light, whereas *GhSACs* respond to light differently. In addition, plant hormone-responsive elements were enriched in the upstream promoter regions of *GhSAC* genes in group I and group III. ABA-responsive elements and MeJA-responsive elements were the most abundant cis-acting hormone responsive elements in the promoters of *GhSAC* genes. This indicated that *GhSAC* genes in group I and group III may be sensitive to ABA and JA than genes in group II.

Gene expression analysis suggests that different *GhSACs* may play specific roles in particular organs or tissues. It is apparent that all *GhSAC* genes are none expressed in leaves and four genes (*GhSAC2.1A/GhSAC2.1D/GhSAC4.2A/GhSAC4.2D*) are predominantly expressed in flowers, suggesting that these proteins may play mainly roles in flowers. The other *GhSAC* genes showed overlapping expression profiles, and expressed in different organs and tissues expect leaves without apparently differences. Further investigation on each GhSAC protein in distinct organs and tissues that will benefit to understand GhSAC proteins in plants while their growth and development. Previous reports about SACs in Arabidopsis show that AtSAC6 protein may play a role mainly in flowers. However, it is intriguing to discover that these four genes belong to subgroup I rather sac6 belongs to subgroup II. Yet GhSACs in subgroup II did not exhibit differential expression patterns of different organs or tissues.

Conclusion

By genome wide analysis of SAC-domain containing genes in *G. hirsutum*, 24 *GhSAC* genes were identified. The *GhSAC* proteins were classified into three different subgroups and showed clear orthologous relationships of SAC members of Arabidopsis, *G. arboreum* and *G. raimondii*. Our expression analysis shows that *GhSAC2.2A, GhSAC2.2D, GhSAC4.2A* and *GhSAC4.2D* are predominantly expressed in flowers. These proteins may play a role mainly in flowers. The present genomic and bioinformatics analyses of GhSAC genesis study provide a solid foundation for further investigation of the cellular functions of *GhSAC* genes.

Methods

Identification of SAC domain- containing proteins

Firstly, We downloaded the HMM profiles of the SAC domains (PF02383) in the Pfam database (http://pfam.xfam.org/) and used it to search the genome database of *G. hirsutum*(ZJU, version 2.1)(http://cotton.zju.edu.cn/), *G. raimondii*(JGI, version 2.0), *G. arboretum*(CRI, version 1.0) (http://www.cottongen.org) using HMMER search program with default E-value, respectively. We also performed HMMER search against *T. cacao* and *V. vinifera* genome databases downloaded from Phytozome database (https://phytozome.jgi.doe.gov) to identify SAC proteins. Previous result has shown that nine SAC proteins exist in *Arabidopsis* [42]. Secondly, *Arabidopsis* SAC domain- containing protein sequences were downloaded from TAIR (http://www.arabidopsis.org/) to use as query to perform the BLASTP against *T. cacao, V. vinifera, G. hirsutum, G. raimondii* and *G. arboreum* genome, respectively. Then, all these sequences were submitted and checked by Interpro (http://www.ebi.ac.uk/interpro) to exclude the sequences without complete SAC-domain.

Chromosomal location, synteny and phylogenetic analysis of SACs

All the *GhSAC* genes were mapped to the *G. hirsutum* genome chromosomes according to approximate position information. MCScanX software (http://chibba.pgml.uga.edu/mcscan2/) was used to do synteny analysis between *GhSAC* genes and *GrSAC* genes and *GaSAC* genes. The local blast + software was used to perform the BLASTP analysis between *G. hirsutum* and *G. raimondii* and *G. arboreum* with the e-value under $1e^{-5}$. The position of SAC domain-containing genes and the blast output were imported into MCScanX and the Dual Systeny Plotter software to exhibit the synteny relationship. Multiple sequence alignment of SAC domain-containing protein sequences from *T. cacao, V. vinifera, G. hirsutum, G. raimondii, G. arboretum* and *Arabidopsis thaliana* were performed using MEGA X with the default parameters. A phylogenetic tree of deduced amino-acid sequences was constructed using the maximum likelihood (ML) method in MEGA X.

Gene Structure And Conserved Motifs Analysis

Structural information on the SAC genes, including chromosomal location and gene length, were obtained from the Phytozome, Cotton Omics Database and CottonGen databases. The domain structures were created by IBS software (version v1.0) and sequence logos were created using Weblogo online software (http://weblogo.threeplusone.com/). The exon/intron structure of each *GhSAC* gene was displayed in Gene Structure Display Server program (http://gsds.cbi.pku.edu.cn/index.php) by comparing the coding sequence and genomic sequence. The conserved motifs prediction was performed using the MEME (http://meme-suite.org/) online program with the following parameters: number of unique motifs: 20; and maximum and minimum search widths: 50 and 6, respectively.

Retrieval And Analysis Of Promoter Sequences

The *G. hirsutum* genome sequences were used to retrieve the promoter sequences (2 kb upstream of the start codon) of the *GhSAC* genes. The analysis of the *GhSAC* promoters was carried out using the Plant- CARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [53].

Expression Profiles of GhSAC genes

The expression levels of *GhSAC* genes containing in different organs or tissues and under different stresses (cold, heat, salt and drought), which were downloaded from the Cotton Omics Database (COD) (http://cotton.zju.edu.cn/).

Quantitative RT-PCR (qRT-PCR) for GhSAC genes

Total RNA was isolated from various tissues using an EASYspin Plus Plant RNA Kit (Aidlab). cDNA was synthesized by using an PrimeScript™ RT reagent Kit with gDNA Eraser (Takara). Cotton ACTIN14 (GenBank accession number: AY305733) was used as an internal control in the PCR assays. The primers were designed based on unique sequences in the *GhSAC* cDNAs, and their sequences are as follows: GhSAC2.1, 5′-CGTTATAATGAGAATGCTAGGCC-3′ and 5′-CCTGCAAGACATTCTGGAATAA-3′; GhSAC4.2, 5′-CAAATCAGCATTACGGGTCAT-3′ and 5′-ATTGTCAGATCCAAGGGAGC-3′; GhSAC7.1, 5′-CGACAAGGGTGAGAAAATGAAA-3′ and 5′-CAAGATTTGTGTGTGAGCTAATGG-3′; GhSAC9.2, 5′-TCTGATTCCTCTGCGTTGC-3′ and 5′-CCAACTTGGTTAGACAAGCCAT-3′. The qRT-PCR was completed with three biological replicates, each comprising four technical replicates. The relative gene expression levels were calculated based on the $2^{-\Delta CT}$ method.

Conflicting of Interests

The authors declare no conflict of interest.

Abbreviations

qRT-PCR: Quantitative real-time polymerase chain reaction; *G.hirsutum*: *Gossypium hirsutumL.* (AADD); *T. cacao*: *Theobroma cacao L.*; *V. vinifera*: *Vitis vinifera L.*; *G.raimondii*: *Gossypium raimondii L.*; *G.arboreum*: *Gossypium arboreum L.* SAC: Suppressor of actin

Declarations

Authors' contributions

XL, WC, SZ and YZ conceived and designed the experiments. XL, YL, JY performed the experiments and analyzed the data. XL drafted the manuscript. XL, SZ and YZ revised the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

The study was supported in part by National Science Foundation in China (31871680) and Agricultural Science and Technology Innovation Program of Chinese Academy of Agricultural Sciences.

Auther details:

[1] School of Life Sciences, Zhengzhou University Zhengzhou, 450052, China

[2] State Key Laboratory of Cotton Biology, Institute of CottonResearch, Chinese Academy of Agricultural Sciences, Anyang 455000, China

References

1. Hughes WE, Cooke FT, Parker PJ (2000a) Sac phosphatase domain proteins. Biochem J 350: 337–352

2. Toker A (1998) The synthesis and cellular roles of phosphatidylinositol 4,5-bisphosphate. Curr Opin Cell Biol 10: 254–261
3. De Camilli, P., Emr, S. D., McPherson, P. S. and Novick, P. (1996) Phosphoinositides as regulators in membrane traffic. Science 271, 1533-1539
4. Corvera, S., D'Arrigo, A. and Stenmark, H. (1999) Phosphoinositides in membrane traffic. Curr. Opin. Cell Biol. 11, 460-465
5. Martin, T. F. (1997) Phosphoinositides as spatial regulators of membrane traffic. Curr. Opin. Neurobiol. 7, 331-338
6. Machesky, L. M. and Insall, R. H. (1999) Signaling to actin dynamics. J. Cell Biol. 146, 267-272
7. Yin, H. L. and Stull, J. T. (1999) Proteins that regulate dynamic actin remodeling in response to membrane signaling. J. Biol. Chem. 274, 32529-32530
8. Takenawa T, Itoh T (2001) Phosphoinositides, key molecules for regulation of actin cytoskeletal organization and membrane traffic from the plasma membrane. Biochim Biophys Acta 1533: 190–206
9. Cremona, O and Di Paolo (1999) Essential Role of Phosphoinositide Metabolism in Synaptic Vesicle Recycling Cell 99, 179–188
10. Higgs, H. N. and Pollard, T. D. (2000) Activation by Cdc42 and Pip2 of Wiskott-Aldrich Syndrome Protein (Wasp) Stimulates Actin Nucleation by Arp2/3 Complex J. Cell Biol. 150, 1311–1320
11. Rohatgi, R., Ho, H. Y., and Kirshner, M. W. (2000) Mechanism of N-WASP activation by CDC42 and phosphatidylinositol 4, 5-bisphosphate J. Cell Biol. 150, 1299–1310
12. Eves, E. M. and Xiong, W. (1998) Akt, a target of phosphatidylinositol 3-kinase, inhibits apoptosis in a differentiating neuronal cell line Mol. Cell. Biol. 18, 2143–2152
13. Novick P, Osmond BC, Botstein D (1989) Suppressors of yeast actin mutations. Genetics 121: 659–674
14. Cleves AE, Novick PJ, Bankaitis VA (1989) Mutations in the *SAC1* gene suppress defects in yeast Golgi and yeast actin function. J Cell Biol 109: 2939–2950
15. Chung, J. K., Sekiya, F., Kang, H. S., Lee, C., Han, J. S., Kim, S. R., Bae, Y. S., Morris, A. J. and Rhee, S. G. (1997) Synaptojanin inhibition of phospholipase D activity by hydrolysis of phosphatidylinositol 4,5-bisphosphate. J. Biol. Chem. 272, 15980-15985
16. Guo, S., Stolz, L. E., Lemrow, S. M. and York, J. D. (1999) SAC1-like domains of yeast SAC1, INP52, and INP53 and of human synaptojanin encode
17. Kearns, B. G., McGee, T. P., Mayinger, P., Gedvilaite, A., Philips, S. E., Kagiwada, S. and Bankaitis, V. A. (1997) Essential role for diacylglycerol in protein transport from the yeast Golgi complex. Nature (London) 387, 101-104
18. Whitters EA, Cleves AE, McGee TP, Skinner HB, Bankaitis VA (1993) SAC1p is an integral membrane protein that influences the cellular requirement for phospholipid transfer protein function and inositol in yeast. J Cell Biol 122: 79–94
19. Foti M, Audhya A, Emr SD (2001) Sac1 lipid phosphatase and Stt4 phosphatidylinositol 4-kinase regulate a pool of phosphatidylinositol 4-phosphate that functions in the control of the actin cytoskeleton and vacuole morphology. Mol Biol Cell 12: 2396–2411
20. Fauman, E. B., Cogswell, J. P., Lovejoy, B., Rocque, W. J., Holmes, W., Montana, V. G., Piwnica-Worms, H., Rink, M. J. and Saper, M. A. (1998) Crystal structure of the catalytic domain of the human cell cycle control phosphatase, Cdc25A. Cell 93, 617-625
21. Kochendorfer, K. U., Then, A. R., Kearns, B. G., Bankaitis, V. A. and Mayinger, P. (1999) Sac1p plays a crucial role in microsomal ATP transport, which is distinct from its function in Golgi phospholipid metabolism. EMBO J. 18, 1506-1515
22. Boyum, R. and Guidotti, G. (1997) Sac1p of Saccharomyces cerevisiae is not involved in ATP release to the extracellular fluid. Biochem. Biophys. Res. Commun. 236, 50-53
23. Salama, N. R., Yeung, T. and Schekman, R. W. (1993) The Sec13p complex and reconstitution of vesicle budding from the ER with purified cytosolic proteins. EMBO J. 12, 4073-4082
24. Barlowe, C, Orci, L., Yeung, T., Hosobuchi, M., Hamamoto, S., Salama, N., Rexach, M. F., Ravazzola, M., Amherdt, M. and Schekman, R. (1994) COPII : a membrane coat formed by Sec proteins that drive vesicle budding from the endoplasmic reticulum. Cell 77, 895-907
25. Matsuoka, K., Orci, L., Amherdt, M., Bednarek, S. Y., Hamamoto, S., Schekman, R. and Yeung, T. (1998) COPII-coated vesicle formation reconstituted with purified coat proteins and chemically defined liposomes. Cell 93, 263-275
26. Jason E. Duex and Johnathan J. Nau Phosphoinositide 5-Phosphatase Fig4p Is Required for both Acute Rise and Subsequent Fall in Stress-Induced Phosphatidylinositol 3,5-Bisphosphate Levels EUKARYOTIC CELL, Apr. 2006, p. 723–731
27. Erdman S, Lin L, Malczynski M, Snyder M (1998) Pheromone-regulated genes required for yeast mating differentiation. J Cell Biol 140: 461–483
28. Luo, W. and Chang, A. (1997) Novel genes involved in endosomal traffic in yeast revealed by suppression of a targeting-defective plasma membrane ATPase mutant. J. Cell Biol. 138, 731-746

29. Nemoto Y, Kearns BG, Wenk MR, Chen H, Mori K, Alb JG, Camilli PD, Bankaitis VA (2000) Functional characterization of a mammalian Sac1 and mutants exhibiting substrate-specific defects in phosphoinositide phosphatase activity. J Biol Chem 275: 34293–34305
30. Minagawa T, Ijuin T, Mochizuki Y, Takenawa T (2001) Identification and characterization of a Sac domain-containing phosphoinositide 5-phosphatase. J Biol Chem 276: 22011–22015
31. Dove SK, Cooke FT, Douglas M, Sayers L, Parker PJ, and Michell RH. 1997 Osmotic Stress activates phosphatidylinositol-3,5-bisphosphate synthesis. Nature 390: 187-192
32. Mueller-Roeber B and Pical C. 2002. Inositol Phospholipid Metabolism in Arabidopsis. Characterized and Putative Isoforms of Inositol Phospholipid Kinase and Phosphoinositide-Specific Phospholipase C11. Plant Physi. 130: 22-46
33. Matsuoka K, Bassham DC, Raikhel NV, Nakamura K (1995) Different sensitivity to wortmannin of two vacuolar sorting signals indicates the presence of distinct sorting machineries in tobacco cells. J Cell Biol 130: 1307–1318
34. Kim DH, Eu Y-J, Yoo CM, Kim Y-W, Pih KT, Jin JB, Kim SJ, Stenmark H, Hwang I (2001) Trafficking of phosphatidylinositol 3-phosphate from the trans-Golgi network to the lumen of the central vacuole in plant cells. Plant Cell 13: 287–301
35. Kost B, Lemichez E, Spielhofer P, Hong Y, Tolias K, Carpenter C, Chua N-H (1999) Rac homologues and compartmentalized phosphatidylinositol 4,5-bisphosphate act in a common pathway to regulate polar pollen tube growth. J Cell Biol 145: 317–330
36. Mikami K, Katagiri T, Luchi S, Yamaguchi-Shinozaki K, Shinozaki K (1998) A gene encoding phosphatidylinositol-4-phosphate 5-kinase is induced by water stress and abscisic acid in Arabidopsis thaliana. Plant J 15: 563–568
37. Meijer HJG, Divecha N, van den Ende H, Musgrave A, Munnik T (1999) Hyperosmotic stress induces rapid synthesis of phosphatidyl-d-inositol 3,5-bisphosphate in plant cells. Planta 208: 294–298
38. Meijer HJ, Berrie CP, Iurisci C, Divecha N, Musgrave A, Munnik T (2001) Identification of a new polyphosphoinositide in plants, phosphatidylinositol 5-monophosphate (PtdIns5P), and its accumulation upon osmotic stress. Biochem J 360: 491–498
39. Pical C, Westergren T, Dove SK, Larsson C, Sommarin M (1999) Salinity and hyperosmotic stress induce rapid increases in phosphatidylinositol 4,5-bisphosphate, diacyglycerol pyrophosphate, and phosphatidylcholine in Arabidopsis thaliana cells. J Biol Chem 274: 38232–38240
40. DeWald DB, Torabinejad J, Jones CA, Shope JC, Cangelosi AR, Thompson JE, Prestwich GD, Hama H (2001) Rapid accumulation of phosphatidylinositol 4,5-bisphosphate and inositol 1,4,5-triphosphate correlates with calcium mobilization in salt-stressed Arabidopsis. Plant Physiol 126: 759–769
41. Zhong R, Burk DH, Nairn CJ, Wood-Jones A, Morrison WH, 3rd, Ye ZH (2005) Mutation of SAC1, an Arabidopsis SAC domain phosphoinositide phosphatase, causes alterations in cell morphogenesis, cell wall synthesis, and actin organization. Plant Cell 17: 1449-1466
42. Ruiqin Zhong and Zheng-Hua Ye (2003) The SAC Domain-Containing Protein Gene Family in Arabidopsis. Plant Physiology 132:544-555
43. Thole JM, Vermeer JE, Zhang Y, Gadella TW, Jr., Nielsen E (2008) Root hair defective4 encodes a phosphatidylinositol-4-phosphate phosphatase required for proper root hair development in Arabidopsis thaliana. Plant Cell 20: 381-395
44. Liu, Z., Ge, X., Yang, Z. et al. Genome-wide identification and characterization of SnRK2 gene family in cotton (*Gossypium hirsutum L.*). BMC Genet 18, 54 (2017).
45. Senchina DS, Alvarez I, Cronn RC, Liu B, Rong J, Noyes RD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in Gossypium. Mol. Biol. Evol. 20, 633–643.
46. Ilsley JL, Sudol M, Winder SJ (2002) The WW domain: linking cell signaling to the membrane cytoskeleton. Cell Signal 14: 183–189
47. Bork P, Sudol M (1994) The WW domain: a signaling site in dystrophin? Trends Biochem Sci 19(12):531-3
48. André B, Springael JY (1994). WWP, a new amino acid motif present in single or multiple copies in various proteins including dystrophin and the SH3-binding Yes-associated protein YAP65. Biochem Biophys Res Commun 205(2):1201-5
49. Sudol M, Chen HI, Bougeret C, Einbond A, Bork P(1995).Characterization of a novel protein-binding module--the WW domain. FEBS Lett 369(1):67-71
50. Changsong Zou, Cairui Lu, Haihong Shang(2013). Genome-Wide Analysis of the Sus Gene Family in Cotton. Journal of Integrative Plant Biology 55 (7): 643–653
51. Hughes WE, Woscholski R, Cooke FT, Patrick RS, Dove SK, McDonald NQ, Parker PJ (2000b) SAC1 encodes a regulated lipid phosphoinositide phosphatase, defects in which can be suppressed by the homologous Inp52p and Inp53p phosphatases. J Biol Chem 275: 801–808
52. Stephen J, Fei N, Juergen E, Zhang S, Dong W, Xue T, Zheng C, Yuan Z. Genome-wide and expression analysis of protein phosphatase 2C in rice and Arabidopsis. BMC Genomics. 2008;9(1):1–21.

53. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P, Rombauts S. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res. 2002;30(1):325.

Figures

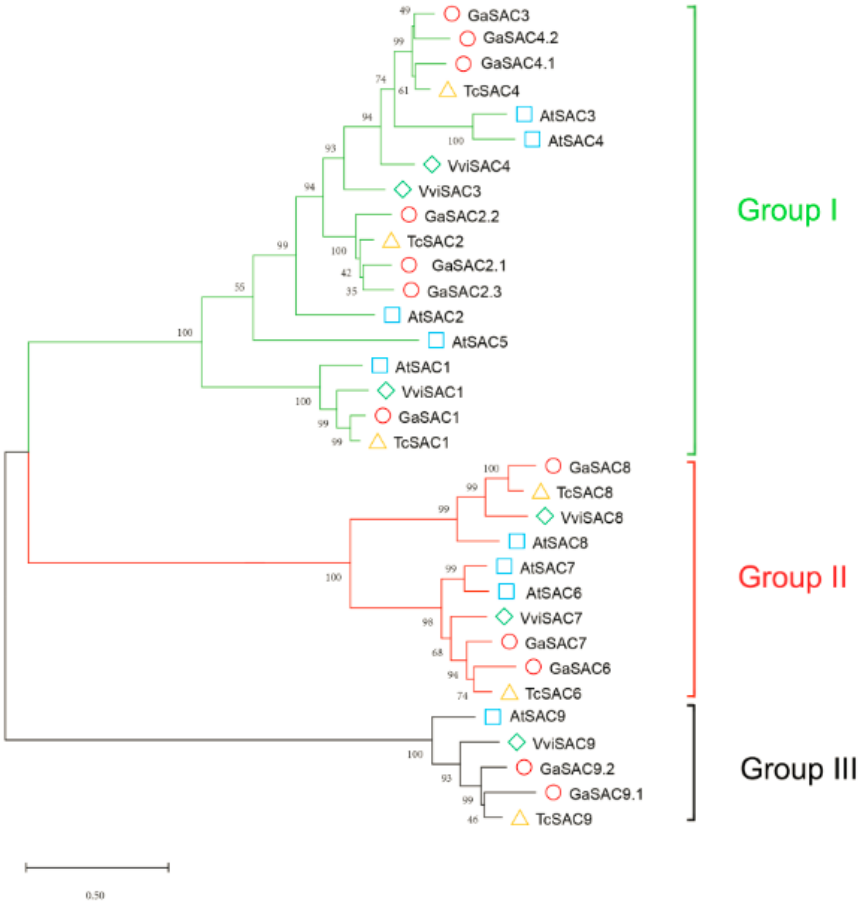

Figure 1

Phylogenetic analysis of SAC proteins from Arabidopsis, G. arboretum, T. cacao and V. vinifera. Full-length amino acid sequences were aligned and phylogenetic tree was constructed by maximum likelihood (ML) method in MEGA X software. The reliability of the internal branches of the tree was evaluated by a bootstrap test (1000 replicates), and the percentages are shown next to the branches. SAC proteins were divided into three distinct groups.

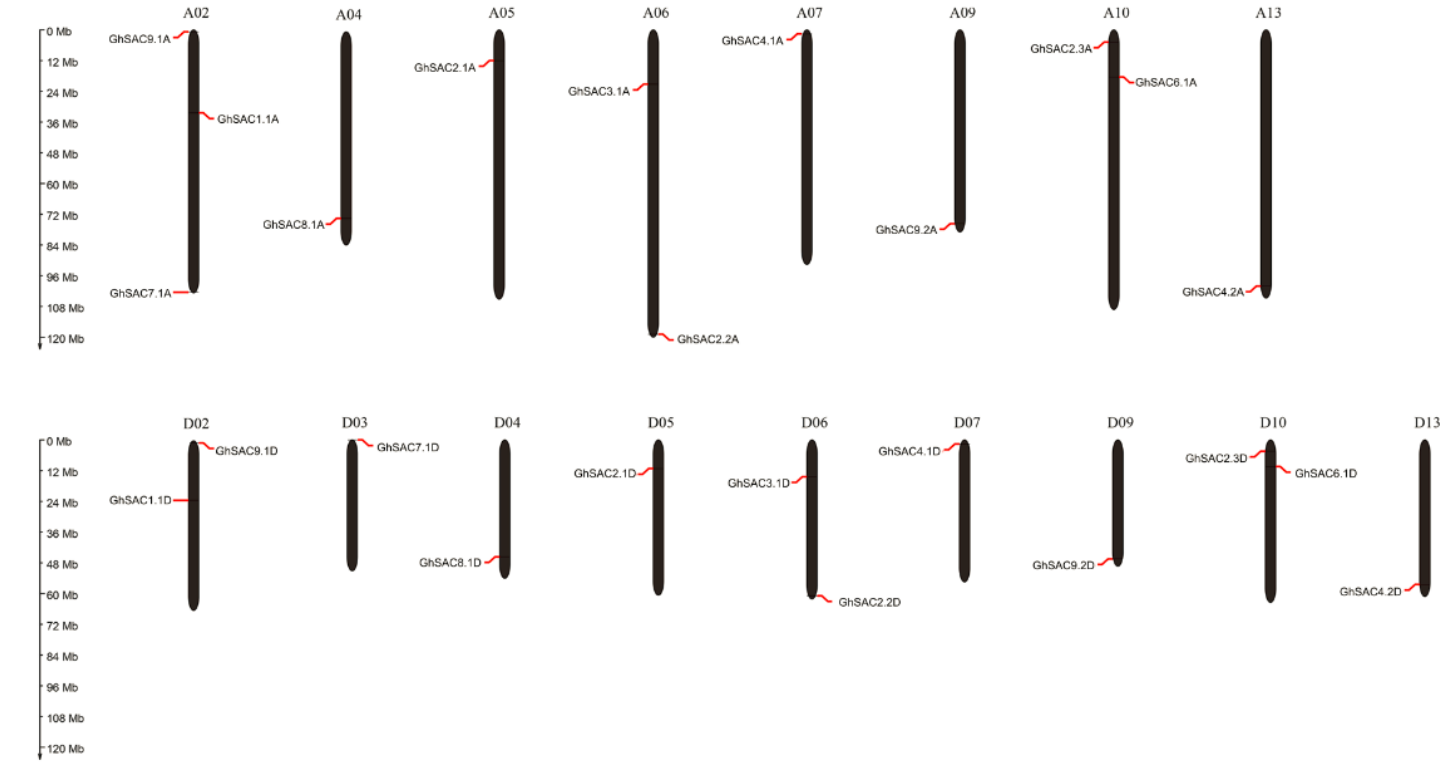

Figure 2

Chromosomal localization of GhSAC genes on Gossypium hirsutum chromosomes.

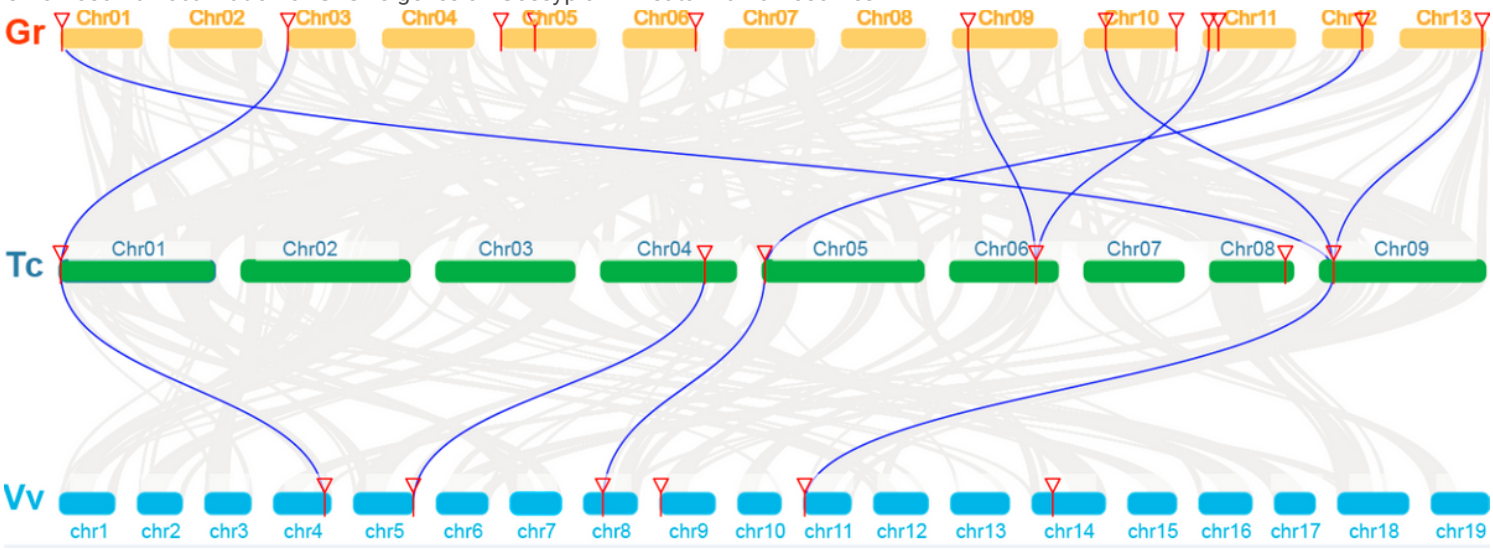

Figure 3

Synteny analysis of SAC genes between G. raimondii and T. cacao and V. vinifera. Gray lines in the background indicate the collinear blocks within these plant genomes, while the blue lines highlight the syntenic SAC gene pairs. The specie names with the prefixes ‘Gr’, ‘Tc’ and ‘Vv’ indicate Gossypium raimondii, Theobroma cacao and Vitis vinifera. Blue lines indicate the syntenic SAC gene pairs between G. hirsutum and G. arboretum while green lines show the gene pairs between G. hirsutum and G. raimondii.

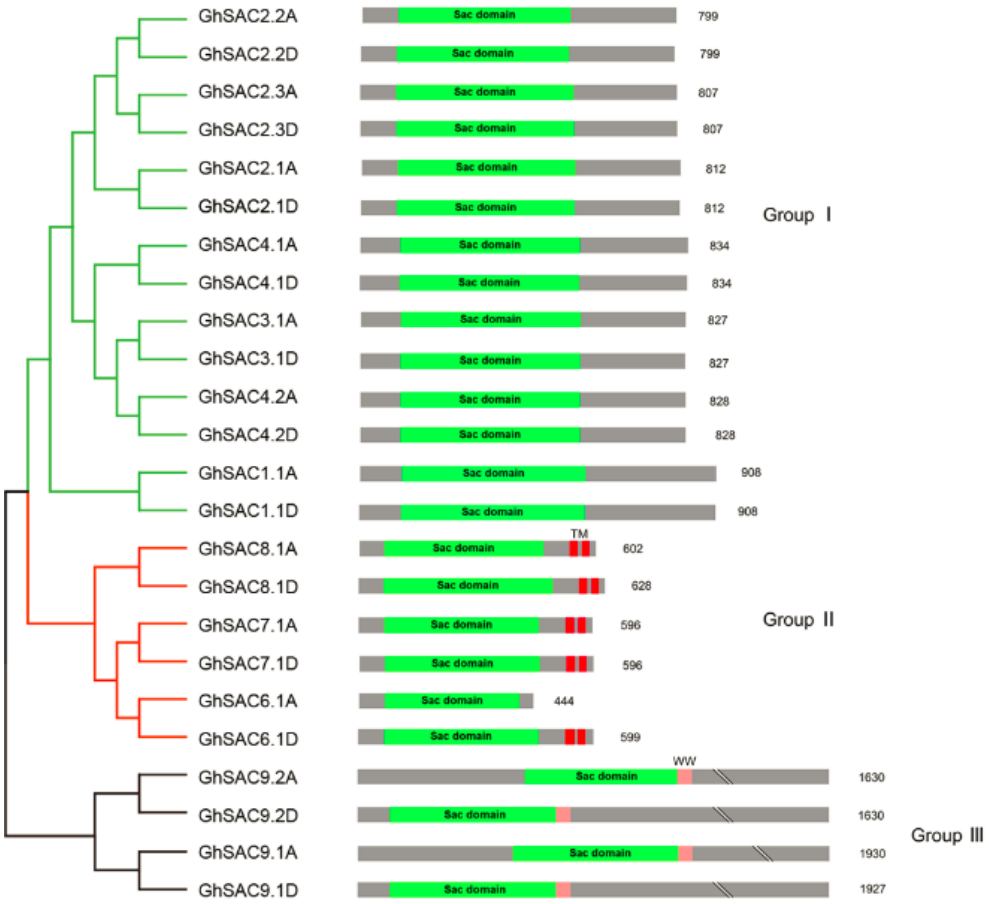

Figure 4

The domain structures of GhSACs. These 24 genes can be sub-divided into three groups based on the result: 1) Group I only contains the SAC domain. 2) Group II contains two C-terminal transmembrane motifs. 2) Group III contains a WW domain followed by the SAC domain. TM: transmembrane motif; WW: WW domain.

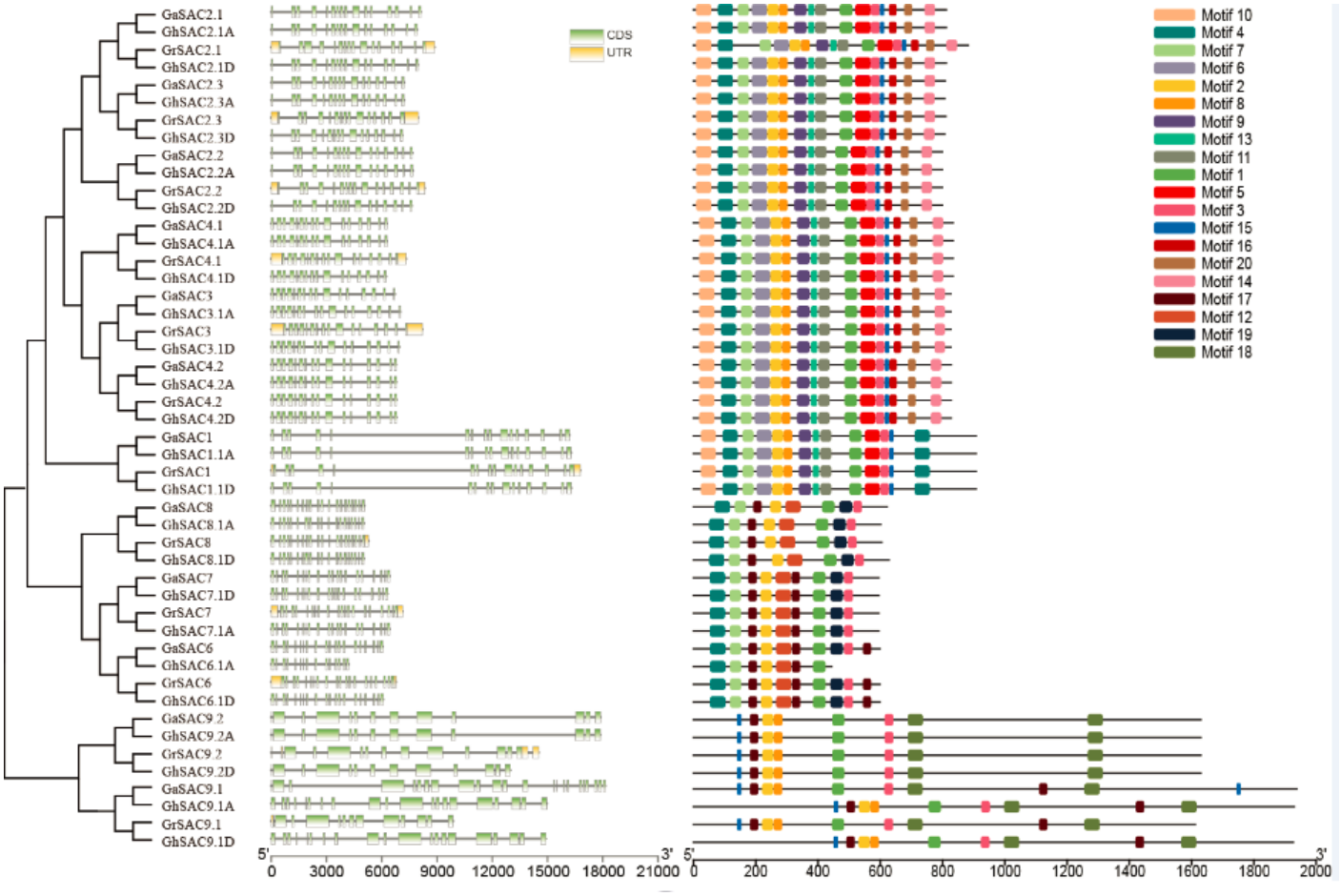

Figure 5

Phylogenetic relationship, exon-intron organizations, and motif analysis of GhSACs. (A) Unrooted phylogenetics tree and structures of GhSAC genes. Unrooted phylogenetic tree was created in MEGA X software with the the maximum likelihood (ML) method. (B) Motif prediction of GhSAC proteins. Twenty motifs were identified by MEME online tool. Each motif is represented by a colored block. The length and position of the motifs could be estimated according to the scale bar.

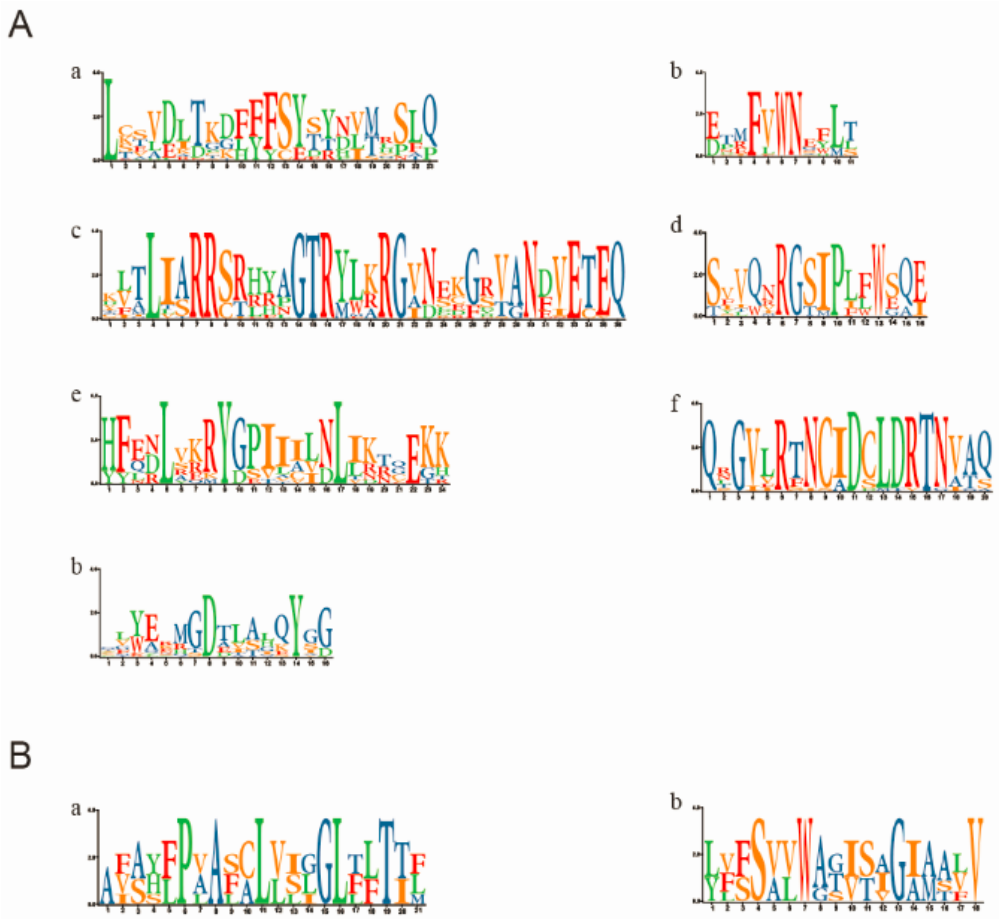

Figure 6

Sequence logos of the G.hirsutum SACs. (A) a ~ f: The seven conserved motifs in the SAC domain. (B) a ~ b: The two transmembrane motifs followed by the SAC domain in Group II.

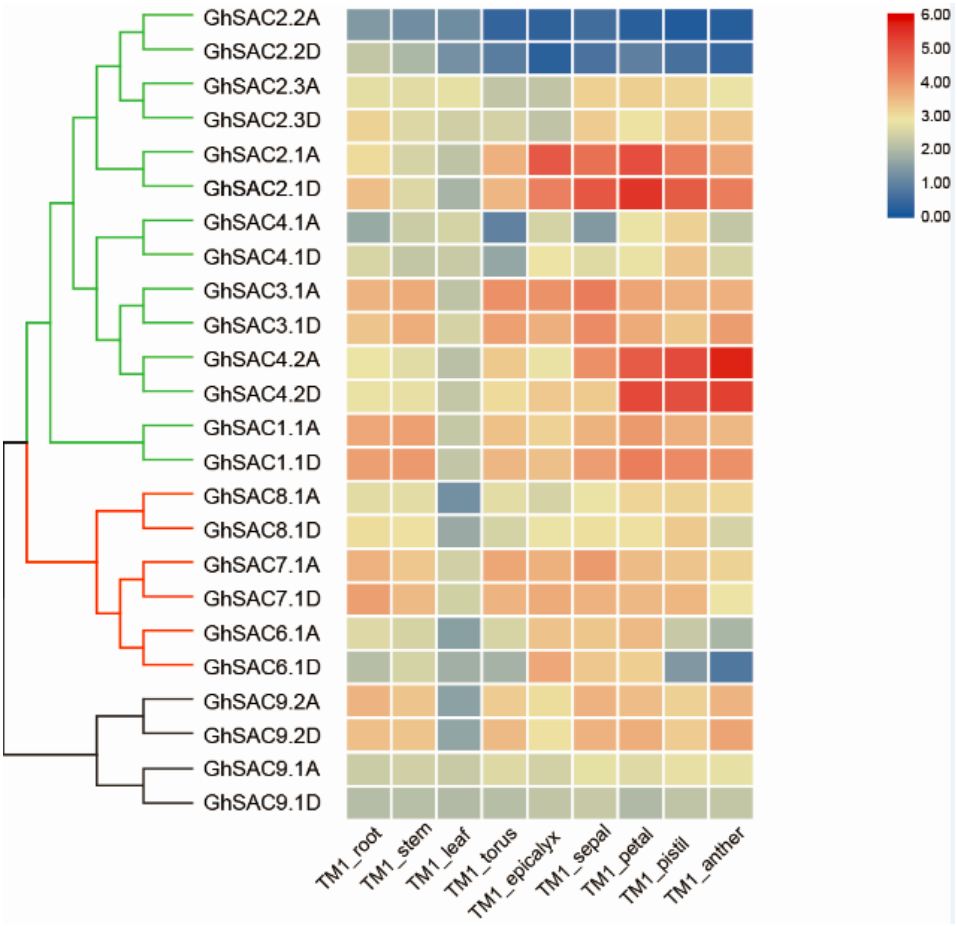

Figure 7

Expression patterns of GhSAC genes in different tissue samples. Color scale bar at the top of map represents log2 transformed FPKM values, which represents low and high expression, respectively. Tissues used for expression profiling are indicated at the bottom of each column. The detail FPKM values are present in Additional file 4.

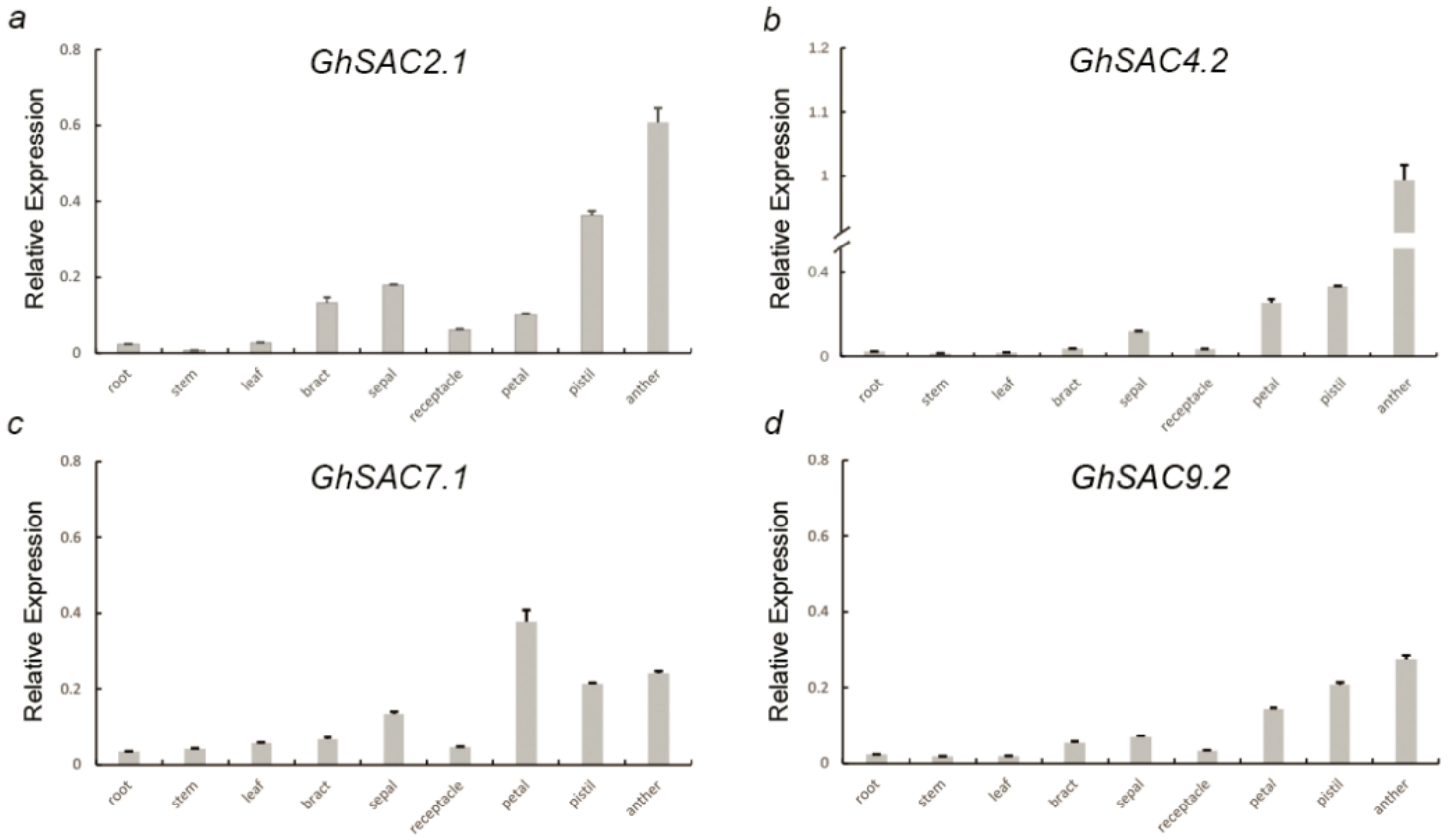

Figure 8

Real-time reverse transcription polymerase chain reaction analysis of four G. hirsutum SAC genes in different tissues.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS4.tif
- TableS1.xlsx
- TableS2.xlsx
- FigureS3.tif
- FigureS2.tif
- FigureS1.tif