BMC Genomics

RESEARCH Open Access

Genome-wide identification of the P4ATPase gene family and its response to biotic and abiotic stress in soybean (*Glycine max* L.)

Jingjing Wei[1], Gaoyang Zhang[1*], Huanhuan Lv[2], Saidi Wang[1], Xingyu Liu[1], Yanli Qi[1], Zhongke Sun[1] and Chengwei Li[1*]

Abstract

Background Soybean is an important legume crop and has significant agricultural and economic value. P4-ATPases (aminophospholipid ATPases, ALAs), one of the classes of P-type ATPases, can transport or flip phospholipids across membranes, creating and maintaining lipid asymmetry and playing crucial roles in plant growth and development. To date, however, the *ALA* gene family and its expression patterns under abiotic and biotic stresses have not been studied in the soybean genome.

Results A total of 27 *GmALA* genes were identified in the soybean genome and these genes were unevenly distributed on 15 chromosomes and classified into five groups based on phylogenetic analysis. The *GmALAs* family had diverse intron–exon patterns and a highly conserved motif distribution. A total of eight domains were found in GmALAs, and all GmALAs had conserved PhoLip_ATPase_C, phosphorylation and transmembrane domains. *Cis*-acting elements in the promoter demonstrated that *GmALAs* are associated with cellular development, phytohormones, environmental stress and photoresponsiveness. Analysis of gene duplication events revealed 24 orthologous gene pairs in soybean and synteny analysis revealed that *GmALAs* had greater collinearity with *AtALAs* than with *OsALAs*. Evolutionary constraint analyses suggested that *GmALAs* have undergone strong selective pressure for purification during the evolution of soybeans. Tissue-specific expression profiles revealed that *GmALAs* were differentially expressed in roots, stems, seeds, flowers, nodules and leaves. The expression pattern of these genes appeared to be diverse in the different developmental tissues. Combined transcriptome and qRT-PCR data confirmed the differential expression of *GmALA*s under abiotic (dehydration, saline, low temperature, ozone, light, wounding and phytohormones) and biotic stresses (aphid, fungi, rhizobia and rust pathogen).

Conclusion In summary, genome-wide identification and evolutionary and expression analyses of the *GmALAs* gene family in soybean were conducted. Our work provides an important theoretical basis for further understanding *GmALAs* in biological functional studies.

Keywords P4-ATPases, Bioinformatics analysis, Expression patterns, Stress response

*Correspondence:
Gaoyang Zhang
gaoyangzhang@haut.edu.cn
Chengwei Li
lcw@haut.edu.cn

[1]School of Biological Engineering, Henan University of Technology, No. 100 Lianhua Street, Zhengzhou High-Tech Development Zone, Zhengzhou 450001, Henan, P. R. China
[2]College of Advanced Interdisciplinary Science and Technology, Henan University of Technology, Zhengzhou 450001, China

Background

Plant P4-ATPases, one of the classes of P-type ATPases, are active membrane transporters that translocate lipids toward the cytosolic side of the biological membranes in eukaryotic cells; these transporters create and maintain lipid asymmetry in the membranes and increase tissue or phenotype polarity [1]. Such lipid flipping also contributes to vesicle formation, vesicle trafficking, cell division or apoptosis and is involved in lipid signal transduction resulting in adaption to environmental stress [2–4]. P4-ATPases proteins are typically composed of polypeptides with four major domains: a nucleotide-binding domain (N-domain), a phosphorylation domain (P-domain), an actuator domain (A-domain) and a β-subunit [5].

P4-ATPases have been studied in protozoan, fungal, plant, and animal species and they contain three major clades of P4 ATPases (P4A, P4B, and P4C) and P4A is predominantly present in plants [4]. The model plant *Arabidopsis thaliana* contains 12 P4 ATPases (ALA1 to ALA12), several of which have been characterized with respect to their transport specificity, subcellular localization, tissue-specific expression and physiological role [1]. *AtALA4* and *AtALA5* are expressed primarily in vegetative tissues and a double knockout of *AtALA4/5* results in plant dwarfism including reduced root hair growth, impaired trichome formation and improperly expanded hypocotyl and leaf epidermal cells [6]. The expression of *AtALA4* is upregulated in the presence of heavy metal ions, and plants lacking *AtALA4* are sensitive to heavy metals [7]. Moreover, *AtALA6* and *AtALA7* are expressed primarily in pollen, and a double knockout of *AtALA6/7* is involved in pollen fertility defects. *AtALA6* overexpression results in increased heat tolerance [6]. The expression pattern of *AtALA3* is similar to that of *AtALA6* and *AtALA7*, which play key roles in pollen tube growth and *Atala3* mutants exhibit reduced pollen tube growth, resulting in fertilization defects, which are worsened by both low and high temperatures [8, 9]. In addition, knockout of *AtALA3* also alters trichome shape and reduces the fitness of *Arabidopsis thaliana* in hot and cold environments [9]. Overexpression of *AtALA2* in *Nicotiana benthamiana* cells results in enlarged multivesicular body structures and *ala2* plants exhibit defects in antiviral defenses [10, 11]. *Arabidopsis* plants lacking *AtALA10* exhibit defects in stomatal opening and closure that likely result from a lack of lysoPC signaling in guard cells [12]. *AtALA10* overexpressing lines are shown to be better at adapting to cold temperatures [13]. The pleiotropic *AtALA1* is involved in stress-related functions, and *ala1* plants are obviously smaller than wild-type plants at chilling temperatures [14]. *GbPATP*, an *AtALA1* homolog in cotton, plays an important role in improving chilling tolerance in cotton and confers cold resistance to *Nicotiana tabacum* [15]. Overexpression of *AtALA1* in maize resulted in enhanced resistance to *F. graminearum* and decreased accumulation of the mycotoxin [16]. In addition, *AtALA1* and *AtALA2* can enhance the immunity of *Arabidopsis thaliana* to viruses [11]. Knockout of *V. dahlia VdDrs2*, *VdNeo1*, *VdP4-4* and *VdDnf1*, which are homologs to cotton P4-ATPases, also effectively decrease the sporulation of *V. dahlia* [17]. In summary, P4 ATPases respond and adapt to several types of biotic and abiotic stresses simultaneously and related mechanisms need to be further studied.

Soybean (*Glycine max* L.), which is rich in protein and oil as well as phytochemicals, is cultivated in several climatic zones. During their life cycle, soybean plants encounter flooding stress, low temperature, drought and salt stresses, virus, bacterial, oomycete, fungal, and nematodes cyst infections, and other stresses, which lead to a reduction in production and quality [13, 18, 19]. The regulatory role of *GmALA* genes in various stress response and plant growth and development has not been assessed in soybean. The investigation of the *GmALA* gene family and its expression patterns under various stresses, including abiotic stress and biotic stress, is highly important for the further study of plant physiology and development in soybean.

In this study, we systematically identified and analyzed the *GmALA* gene family in soybeans and clarified the expression pattern of *GmALAs* in soybean under stress. Phylogenetic relationship, physicochemical property, motif composition, conserved domain, gene structure, duplication event, *cis*-element composition, chromosomal location, and collinearity analyses were performed using the current soybean genome sequence data. Moreover, the expression patterns of *GmALAs* in different tissues and developmental tissues were presented. In addition, the responses of *GmALAs* to various abiotic and biotic treatments were also analyzed. This study provides a valuable foundation for further functional investigations of *GmALAs* and offers new insight into the mechanism of *GmALAs* under abiotic stresses.

Results

Identification of *GmALA* genes in the soybean genome

To identify the GmALA members in the soybean genome, ALAs from *A. thaliana* and *O. sativa* were used as queries. A total of 27 homologous GmALAs were identified in soybean via BLASTP (BLOSUM62) against the Phytozome database and were renamed GmALA1 ~ GmALA27 (Table 1). Information on these *GmALA* genes and their corresponding proteins is shown in Table 1, which includes gene ID, chromosome location, number of exons, protein length (AA), molecular weight (M. W), theoretical isoelectric point (pI), instability index (II), grand average of hydropathicity (GRAVY), aliphatic

Table 1 Information of the *GmALAs* in soybean

Gene ID	Chromosome	Name	Strand	Start (bp)	End (bp)	AA	M.W	PI	II	Aliphatic index	GRAVY	Subcellular localization	Exons
Glyma.13G348200	Chr13	*GmALA1*	-	43,822,574	43,829,935	1217	137348.98	5.89	39.9	90.88	-0.064	plasmolemma	10
Glyma.06G208900	Chr06	*GmALA2*	-	20,515,068	20,521,900	1190	134370.25	5.5	38.22	94.55	-0.011	plasmolemma	11
Glyma.15G025800	Chr15	*GmALA3*	+	2,089,114	2,095,958	1224	138274.43	5.97	40.24	92.82	-0.031	plasmolemma	10
Glyma.04G166100	Chr04	*GmALA4*	+	41,682,491	41,693,023	1166	132536.06	6.52	35.85	91.56	-0.014	plasmolemma	11
Glyma.05G015400	Chr05	*GmALA5*	+	1,413,681	1,420,973	1205	136891.44	5.89	40.64	92.51	-0.046	plasmolemma	11
Glyma.08G268900	Chr08	*GmALA6*	+	34,656,620	34,664,578	1198	135891.86	6.49	38.35	92.49	0.005	plasmolemma	11
Glyma.17G123800	Chr17	*GmALA7*	-	9,879,196	9,885,937	1217	138260.9	6.04	42.5	91.12	-0.082	plasmolemma	11
Glyma.04G144900	Chr04	*GmALA8*	-	26,498,605	26,505,403	1189	134263.24	5.54	38.15	94.96	-0.01	plasmolemma	11
Glyma.18G129200	Chr18	*GmALA9*	+	17,461,040	17,470,275	943	107660.45	7.18	43.36	93.24	0.025	plasmolemma	18
Glyma.06G196400	Chr06	*GmALA10*	-	17,625,990	17,632,800	998	113349.79	5.32	37.3	93.42	0.045	plasmolemma	10
Glyma.04G129200	Chr04	*GmALA11*	-	18,028,813	18,036,924	1202	134510.94	7.28	34.32	97.69	0.034	plasmolemma	7
Glyma.08G190400	Chr08	*GmALA12*	-	15,291,579	15,298,606	1231	139438.52	6.14	40.07	90.39	-0.081	plasmolemma	10
Glyma.07G007700	Chr07	*GmALA13*	-	563,432	570,287	1231	139446.58	6.16	41.81	90.63	-0.081	plasmolemma	10
Glyma.01G092900	Chr01	*GmALA14*	+	28,040,320	28,049,286	1198	135979.38	6.23	37.17	90.12	-0.054	plasmolemma	11
Glyma.06G315100	Chr06	*GmALA15*	+	50,401,945	50,408,153	1226	137182.64	6.18	38.93	97.14	0.004	plasmolemma	7
Glyma.02G129500	Chr02	*GmALA16*	-	13,270,168	13,279,129	1198	136217.61	6.19	39.07	89.96	-0.058	plasmolemma	11
Glyma.16G099700	Chr16	*GmALA17*	-	19,248,629	19,256,736	1173	132832.33	6.17	40.21	93.05	0.012	plasmolemma	11
Glyma.08G293900	Chr08	*GmALA18*	-	40,887,462	40,910,754	1227	139937.23	8.11	43.55	91.19	-0.06	plasmolemma	27
Glyma.09G273900	Chr09	*GmALA19*	+	48,997,462	49,005,285	1296	145793.59	6.09	41.62	92.47	-0.111	plasmolemma	7
Glyma.13G295100	Chr13	*GmALA20*	+	39,436,754	39,451,888	1102	124733.37	6.56	39.2	99.53	0.038	plasmolemma	24
Glyma.08G228800	Chr08	*GmALA21*	+	18,684,920	18,691,455	1180	132259.33	6.48	37.89	97.31	0.023	plasmolemma	7
Glyma.05G006800	Chr05	*GmALA22*	+	643,482	653,987	1194	136409.84	7	38.37	92.72	-0.111	plasmolemma	11
Glyma.12G205800	Chr12	*GmALA23*	-	36,604,017	36,619,778	1107	125392.84	6.28	40.21	97.31	0.008	plasmolemma	24
Glyma.18G213100	Chr18	*GmALA24*	+	49,992,860	50,002,425	1295	145880.62	6.07	40.64	92.3	-0.127	plasmolemma	7
Glyma.15G215400	Chr15	*GmALA25*	+	35,264,851	35,271,438	1180	131835.7	6.1	36.73	98.47	0.049	plasmolemma	7
Glyma.19G007400	Chr19	*GmALA26*	+	739,763	750,443	1189	136090.45	6.73	38.82	93.44	-0.118	plasmolemma	11
Glyma.16G220100	Chr16	*GmALA27*	-	37,724,844	37,732,883	1172	131819.57	6.15	40.02	95.6	-0.038	plasmolemma	7

Abbreviation: AA: amino acids, M.W: Molecular weight, II: instability index, GRAVY: Grand average of hydropathicity

index and subcellular location. The protein length of GmALAs varied greatly from 943 aa (GmALA9) to 1296 aa (GmALA19). The pI ranges from 5.32 (GmALA10) to 8.11(GmALA18), whereas the molecular weight varies from 107.66 kDa (GmALA9) to 145.88 kDa (GmALA24). The II of the proteins ranged from 37.3 (GmALA10) to 43.55 (GmALA18), which predicted that 12 of the GmALAs were unstable because their II value was less than 40 and that the other 15 GmALA proteins were stable. Moreover, the ranges of the aliphatic index and GRAVY ranged from 89.96 (GmALA16) to 99.53 (GmALA20) and from −0.127 (GmALA24) to 0.049 (GmALA25), respectively. A GRAVY value less than 0 predicted that most GmALAs were hydrophilic proteins, except for GmALA9-11, GmALA15, GmALA17, GmALA20, GmALA21, GmALA23 and GmALA25. It seemed that all the *GmALAs* were located in the plasmolemma according to the predicted subcellular localization. Other detailed information about the pI and MW of the GmALAs is presented in Table 1.

To understand the genomic organization and distribution of *GmALAs* on the soybean chromosomes, we constructed a chromosomal map. Our results showed that 27 *GmALAs* were unevenly and irregularly distributed on 15 soybean chromosomes (Fig. 1). Chromosome 08 contained 4 *GmALAs* (*GmALA6*, *GmALA12*, *GmALA18* and *GmALA21*) which was greater than other chromosome, while both chromosomes 04 and 06 had 3 *GmALAs*. Chromosomes 05, 13, 15, 16 and 18 all had two *GmALAs*. In contrast, chromosome 01, 02, 07, 09, 12, 17 and 19 each consisted of one *GmALA*. The remaining chromosomes did not have any homologous genes to *AtALAs* or *OsALAs*.

The chromosomal position of each *GmALA* gene is mapped based on the soybean genome. Chromosome numbers are indicated on the top of each scaffold. The chromosome size is shown by the vertical scale.

Phylogenetic analysis of the *GmALA* gene family

To investigate the phylogenetic relationships of the ALA family between *Arabidopsis*, rice and soybean, a phylogenetic tree was generated based on the full-length protein sequences of 12 AtALAs, 8 OsALAs and 27 GmALAs by using MEGA 7.0 with the neighbor-joining method (1000 bootstrap replications) (Fig. 2). Based on the P4-ATPase family classification standard and the bootstrap values (> 50%) of the phylogenetic tree, the GmALA family was clustered into five groups, and their distribution in each subgroup was rather uneven (Fig. 2). Specifically, Group 1 contained seven GmALAs (GmALA11, 15, 19, 21, 24, 25 and 27), AtALA1 and three OsALAs (OsALA-3). Twelve GmALAs (GmALA2, 4–8, 10, 14, 16, 17, 22 and 26) were clustered with five AtALAs (AtALA8-12) and three OsALAs (OsALA4-6) in Group 2. Group 3 contained four GmALAs (GmALA1, 3, 12 and 13), four AtALAs (AtALA4-7) and OsALA7. Group 4 consisted of only two GmALAs (GmALA9 and 18), AtALA3. Two GmALAs (GmALA20 and 23) were clustered with AtALA2 and OsALA8 in Group 5. Groups 1–4 and 5 belong to the P4A-ATPase and P4C-ATPase subclades, respectively. Moreover, group 1 was further subdivided into the P4Ad group, and groups 2–4 belonged to the P4Ae group (Fig. 2).

A phylogenetic tree was constructed with 12 AtALA protein sequences, 8 OsALA protein sequences and 27 GmALA protein sequences. All ALA protein sequences are downloaded from Phytozome (https://phytozome-next.jgi.doe.gov/). The five subclades of the ALA family are highlighted in distinct colors. The blue triangles, yellow squares and red pentacles represent the ALA genes from

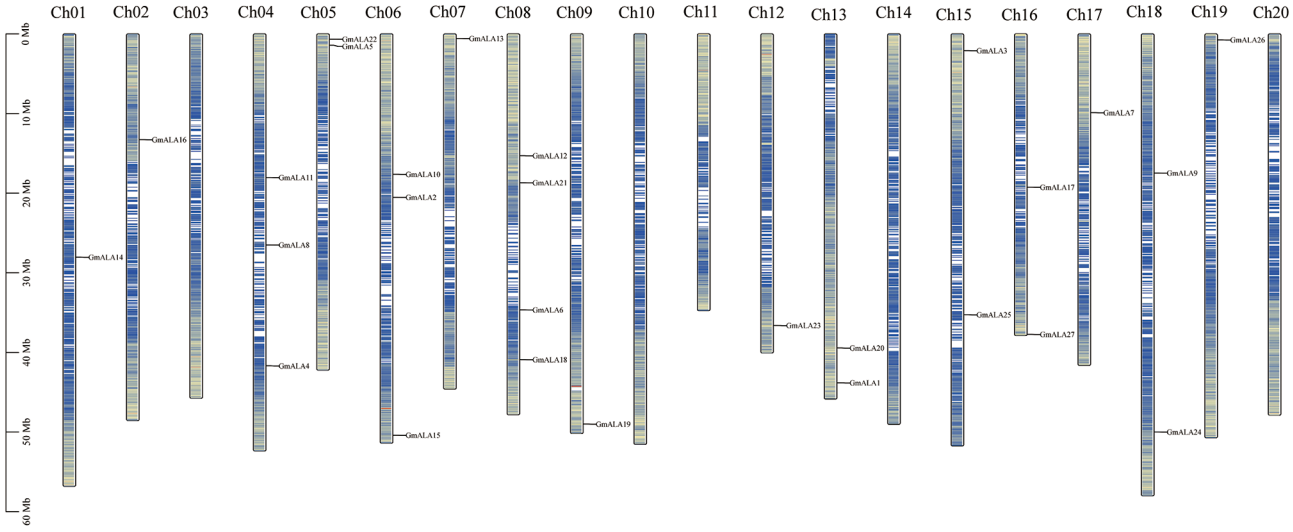


Fig. 1 Chromosomal distribution of *GmALAs* in the soybean genome

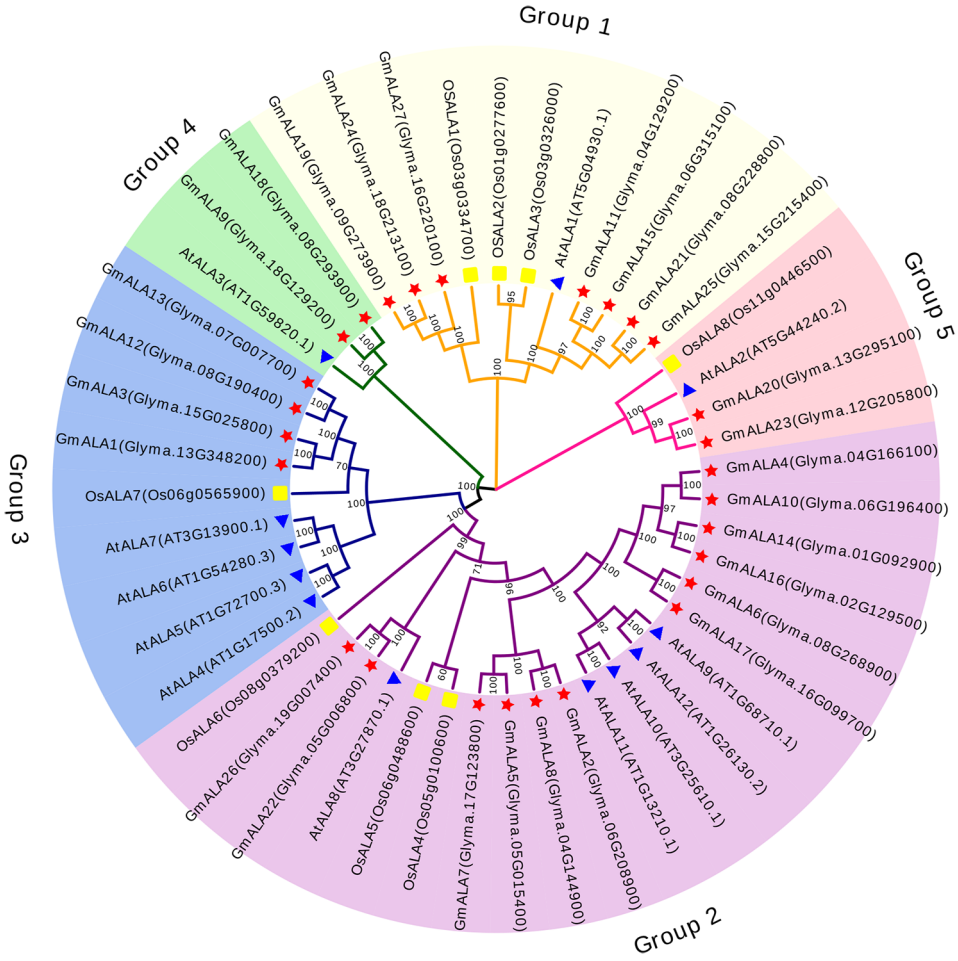


Fig. 2 Phylogenetic tree of the ALA protein sequences from *Arabidopsis thaliana*, rice and soybean

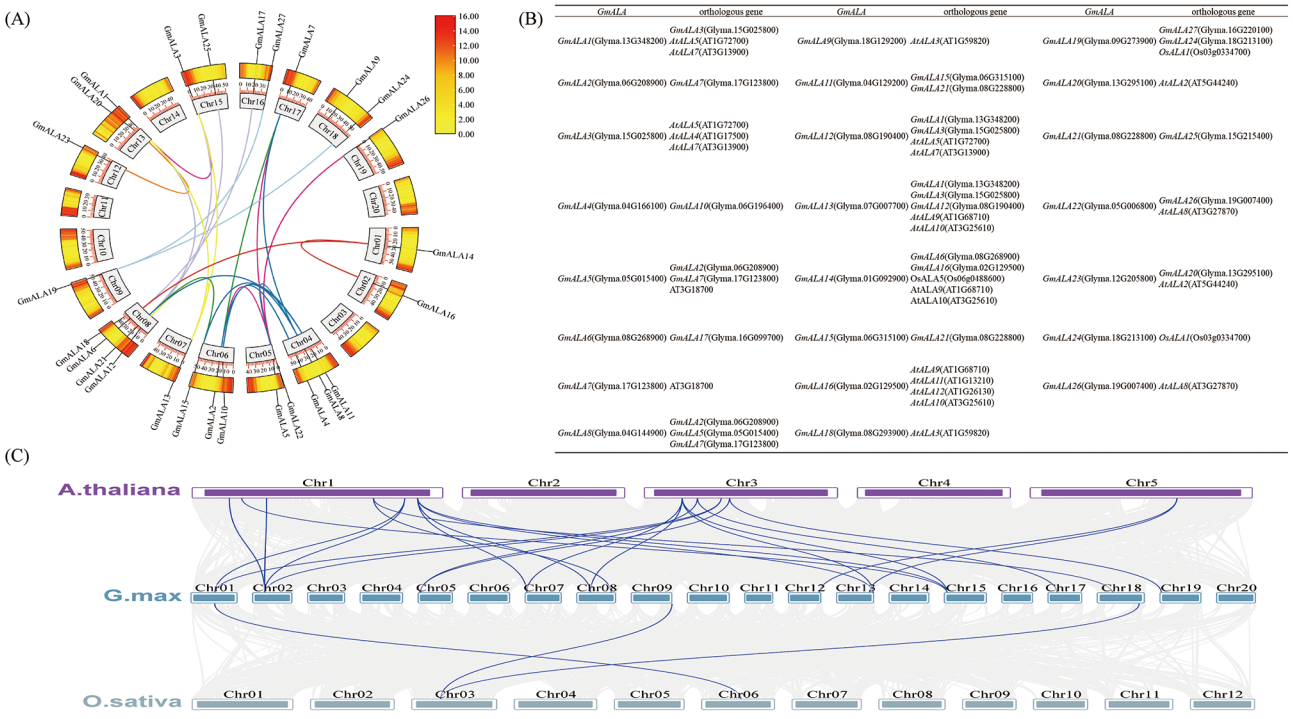
Arabidopsis, rice and soybean, respectively. MEGA 7.0 was used to carry out protein sequence alignment and subsequent bootstrap tree construction with the neighbor-joining (NJ) method and 1000 bootstrap values.

Synteny analysis of *GmALA* genes and calculation of Ka/Ks

Tandem and segmental duplications of the *GmALAs* were obtained by collinear analysis. As a result, 24 pairs of segmentally duplicated genes were identified on fifteen chromosomes (Chr1, Chr2, Chr4-9, Chr12, Chr13, and Chr15-19) (Fig. 3A, B). There were no tandemly duplicated genes on the soybean chromosomes. Syntenic analysis was also carried out for *A. thaliana*, rice and soybean (Fig. 3B, C). The results revealed that 14 (51.8%) *GmALAs* showed pairwise synteny with genes in the *A. thaliana* genome, and 3 (11.1%) *GmALAs* showed pairwise synteny with genes in rice. To investigate whether Darwinian positive selection was involved in the divergence of *GmALAs* after duplication and to trace the dates of the duplication blocks, the ratio of nonsynonymous substitution per site (Ka) to the synonymous substitution rate per site (Ks) was used to show the molecular evolution of *GmALAs*, and the ratios of Ka/Ks of all the tandem amplifications and segmental duplications was showed on Supplementary Table S1. The substitution rate ratios (Ka/Ks) of 17 paralogous pairs were calculated using TBtools. Ks was used to calculate the approximate dates of duplication events. The segmental duplications of the *GmALAs* in the soybean genome was likely originated from 4.59 Mya (million years ago, Ks = 0.0560) to 60.34 Mya (Ks = 0.7362), with a mean value of 25.85 Mya (Ks = 0.3154). In this study, the Ka/Ks ratios of 16 segmental duplication pairs were less than 0.3, and those of only one segmental duplication pair were more than 0.3, which demonstrated the possibility of significant functional divergence of some *GmALAs* after duplication events.

Motif compositions, gene structures, conserved domains and *cis*-acting regulatory elements of GmALAs

Ten conserved motifs (motif1-motif10) existed in GmALA proteins (Fig. 4A). The lengths and motif sequences are shown in Table 2. The length of the 10 motifs ranged from 8 AA (motif 10) to 20 AA (motif



GmALA	orthologous gene	*GmALA*	orthologous gene	*GmALA*	orthologous gene
GmALA1(Glyma.13G348200)	*GmALA3*(Glyma.15G025800) *AtALA5*(AT1G72700) *AtALA7*(AT3G13900)	*GmALA9*(Glyma.18G129200)	*AtALA3*(AT1G59820)	*GmALA19*(Glyma.09G273900)	*GmALA27*(Glyma.16G220100) *GmALA24*(Glyma.18G213100) *OsALA1*(Os03g0334700)
GmALA2(Glyma.06G208900)	*GmALA7*(Glyma.17G123800)	*GmALA11*(Glyma.04G129200)	*GmALA15*(Glyma.06G315100) *GmALA21*(Glyma.08G228800)	*GmALA20*(Glyma.13G295100)	*AtALA2*(AT5G44240)
GmALA3(Glyma.15G025800)	*AtALA5*(AT1G72700) *AtALA4*(AT1G17500) *AtALA7*(AT3G13900)	*GmALA12*(Glyma.08G190400)	*GmALA1*(Glyma.13G348200) *GmALA3*(Glyma.15G025800) *AtALA5*(AT1G72700) *AtALA7*(AT3G13900)	*GmALA21*(Glyma.08G228800)	*GmALA25*(Glyma.15G215400)
GmALA4(Glyma.04G166100)	*GmALA10*(Glyma.06G196400)	*GmALA13*(Glyma.07G007700)	*GmALA1*(Glyma.13G348200) *GmALA3*(Glyma.15G025800) *GmALA12*(Glyma.08G190400) *AtALA9*(AT1G68710) *AtALA10*(AT3G25610)	*GmALA22*(Glyma.05G006800)	*GmALA26*(Glyma.19G007400) *AtALA8*(AT3G27870)
GmALA5(Glyma.05G015400)	*GmALA2*(Glyma.06G208900) *GmALA7*(Glyma.17G123800) AT3G18700	*GmALA14*(Glyma.01G092900)	*GmALA6*(Glyma.08G268900) *GmALA16*(Glyma.02G129500) OsALA5(Os06g0488600) AtALA9(AT1G68710) AtALA10(AT3G25610)	*GmALA23*(Glyma.12G205800)	*GmALA20*(Glyma.13G295100) *AtALA2*(AT5G44240)
GmALA6(Glyma.08G268900)	*GmALA17*(Glyma.16G099700)	*GmALA15*(Glyma.06G315100)	*GmALA21*(Glyma.08G228800)	*GmALA24*(Glyma.18G213100)	*OsALA1*(Os03g0334700)
GmALA7(Glyma.17G123800)	AT3G18700	*GmALA16*(Glyma.02G129500)	*AtALA9*(AT1G68710) *AtALA11*(AT1G13210) *AtALA12*(AT1G26130) *AtALA10*(AT3G25610)	*GmALA26*(Glyma.19G007400)	*AtALA8*(AT3G27870)
GmALA8(Glyma.04G144900)	*GmALA2*(Glyma.06G208900) *GmALA5*(Glyma.05G015400) *GmALA7*(Glyma.17G123800)	*GmALA18*(Glyma.08G293900)	*AtALA3*(AT1G59820)		

Fig. 3 Collinearity analysis of the *ALA* gene family in soybean, rice and *Arabidopsis*. (**A**) Schematic representations of the interchromosomal relationships of *GmALAs*. The colored lines indicate the synteny blocks in the soybean genome. (**B**) *GmALA* orthologous genes in soybean, Arabidopsis and rice. (**C**) Synteny analysis of the *ALA* genes in *Arabidopsis thaliana*, rice and soybean. The gray lines in the background indicate the collinear blocks within *Arabidopsis thaliana*, rice and soybean; the blue lines indicate the syntenic *GmALA* gene pairs

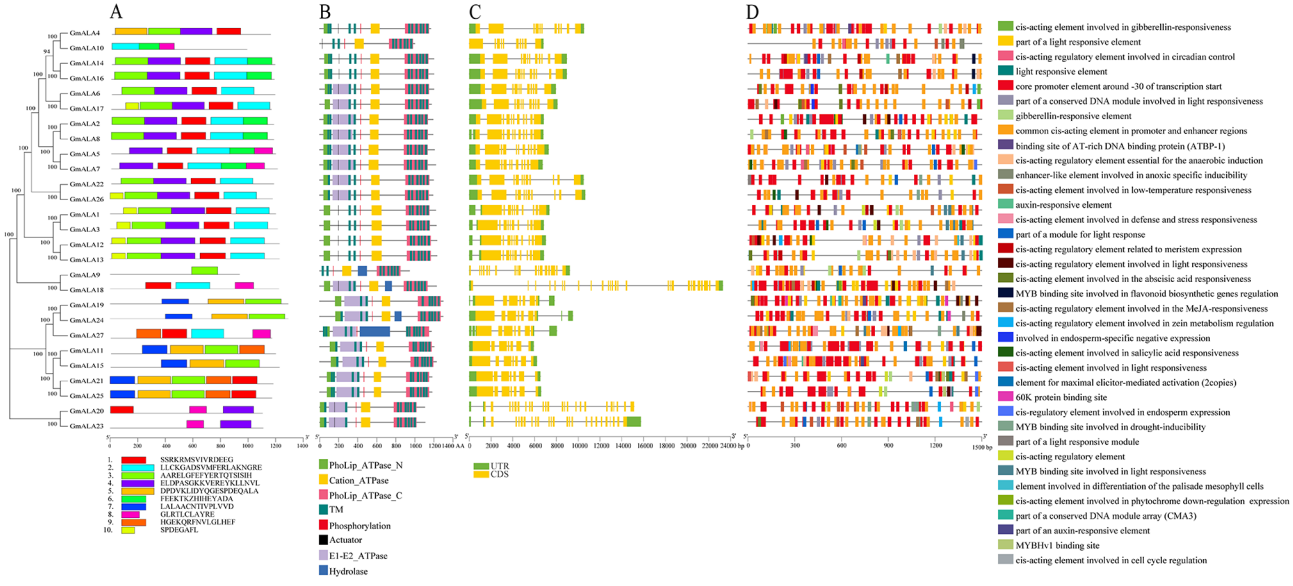


Fig. 4 Analysis of conserved motifs, protein domains, gene structure and *cis*-acting elements of the *GmALAs* promoter in soybean. (**A**) Motif composition of soybean GmALA proteins. The motifs (numbered 1–10) are displayed in differently colored boxes. The lengths of the proteins can be estimated using the scale at the bottom. (**B**) Conserved domains and their distribution of GmALAs in soybean. The conserved domains are named below and are presented in different colors. (**C**) Gene structure of soybean *GmALA* genes. The UTRs, CDSs, and introns are represented by green boxes, yellow boxes and gray lines, respectively. (**D**) *Cis*-acting element analysis of *GmALA* promoters

Table 2 Analysis of the 10 conserved motifs of GmALAs in soybean

Motif	Loge	Best possible match	Width
1		SSRKRMSVIVRDEEG	15
2		LLCKGADSVMFERLAKNGRE	20
3		AARELGFEFYERTQTSISIH	20
4		ELDPASGKKVEREYKLLNVL	20
5		DPDVKLIDYQGESPDEQALA	20
6		FEEKTKZHIHEYADA	15
7		LALAACNTIVPLVVD	15
8		GLRTLCLAYRE	11
9		HGEKQRFNVLGLHEF	15
10		SPDEGAFL	8

2–5). All the GmALAs had motifs 4 except for GmALA9-11, 15, 18–19, 21, 24–25, and 27. However, most of GmALAs contains five motifs, except for GmALA4, 6, 11, 22, 24, and 27 which have four motifs. In addition, GmALA9 had only one motif4 but GmALA23 contained motif 4, 8. And GmALA10, 15, 18–20 showed three motifs. The Pfam program was used to identify eight types of domains-based GmALA proteins sequence (Fig. 4B). As expected, the PhoLip_ATPase_N, Cation_ATPase, PhoLip_ATPase_C, TMs, phosphorylation, actuator, E1-E2_ATPase and hydrolase conserved domains existed in the GmALA proteins and all GmALA proteins contained the PhoLip_ATPase_C conserved domain, TMs, phosphorylation, and PhoLip_ATPase_N domain, except for GmALA9 and 10, which lack the PhoLip_ATPase_N domain. Additionally, GmALAs gene family contains 7–10 TMs and the number difference of TMs between GmALAs mainly occurs at the N-terminal, such as GmALA1 has 2 TMs at the N-terminal and 6 at C-terminal. Some GmALAs (GmALA2, 5–8, 11, 15, and 17–27) had E1-E2_ATPase domains, while all GmALAs, except for GmALA27, contains Cation_ATPase domain. Besides, all GmALAs, except for GmALA9, 11, 15, 19, 21, 24, 25, and 27, had Actuator domains and only GmALA9, 18, 24, and 27 had hydrolase domains. The gene structure of *GmALAs* exhibited diverse exon–intron patterns and the number of exons varied from 7 to 27 (Fig. 4C). For instance, *GmALA18*, *20* and *23* had more than twenty exons and *GmALA18* notably had twenty-seven exons. Meanwhile, the *GmALA* members with high homology, such as *GmALA7* and *GmALA8*, had highly similar exon–intron structures (intron number and exon length).

The *cis*-acting elements in the 1500 bp upstream promoter sequences of the *GmALAs* were analyzed. A total of 37 *cis*-regulatory elements were identified in the promoter regions of *GmALAs*, and these elements were divided into four categories environmental stress-, photoresponsive-, phytohormone-, and plant growth- and development- responsive elements (Fig. 4D). The stress−responsive elements included elements essential for anaerobic induction (ARE), anoxic specific inducibility (GC-motif), low-temperature responsiveness (LTR), drought-inducibility (MBS), light responsiveness (MRE), and defense and stress responsiveness (TC-rich repeats). The photoresponsive element contained Box 4, the GT1-motif, the G-box and the GATA- motif. There were eleven types of *cis*-acting hormone- responsive elements, and several important responsive elements, such as abscisic acid responsiveness elements (ABREs), MeJA-responsive elements (CGTCA motifs and TGACG motifs), salicylic acid-responsive elements (TCA elements), and auxin-responsive elements (TGA boxes), were abundant. The plant growth and development responsive element group included five types of *cis*-acting elements, such as CAT-box (meristem expression), circadian (circadian control), GCN4_motif (endosperm expression), HD-Zip 1 (differentiation of the palisade mesophyll cells) and O_2-site (zein metabolism).

Expression patterns of *GmALAs* in different tissues

The tissue expression profiles demonstrated that all *GmALAs* were detected in tissues with diverse expression patterns (Fig. 5A). For instance, *GmALA18* had the highest expression level in stems, root hairs and shoot apical meristems (sams) tissues. *GmALA6*, *GmALA12*,

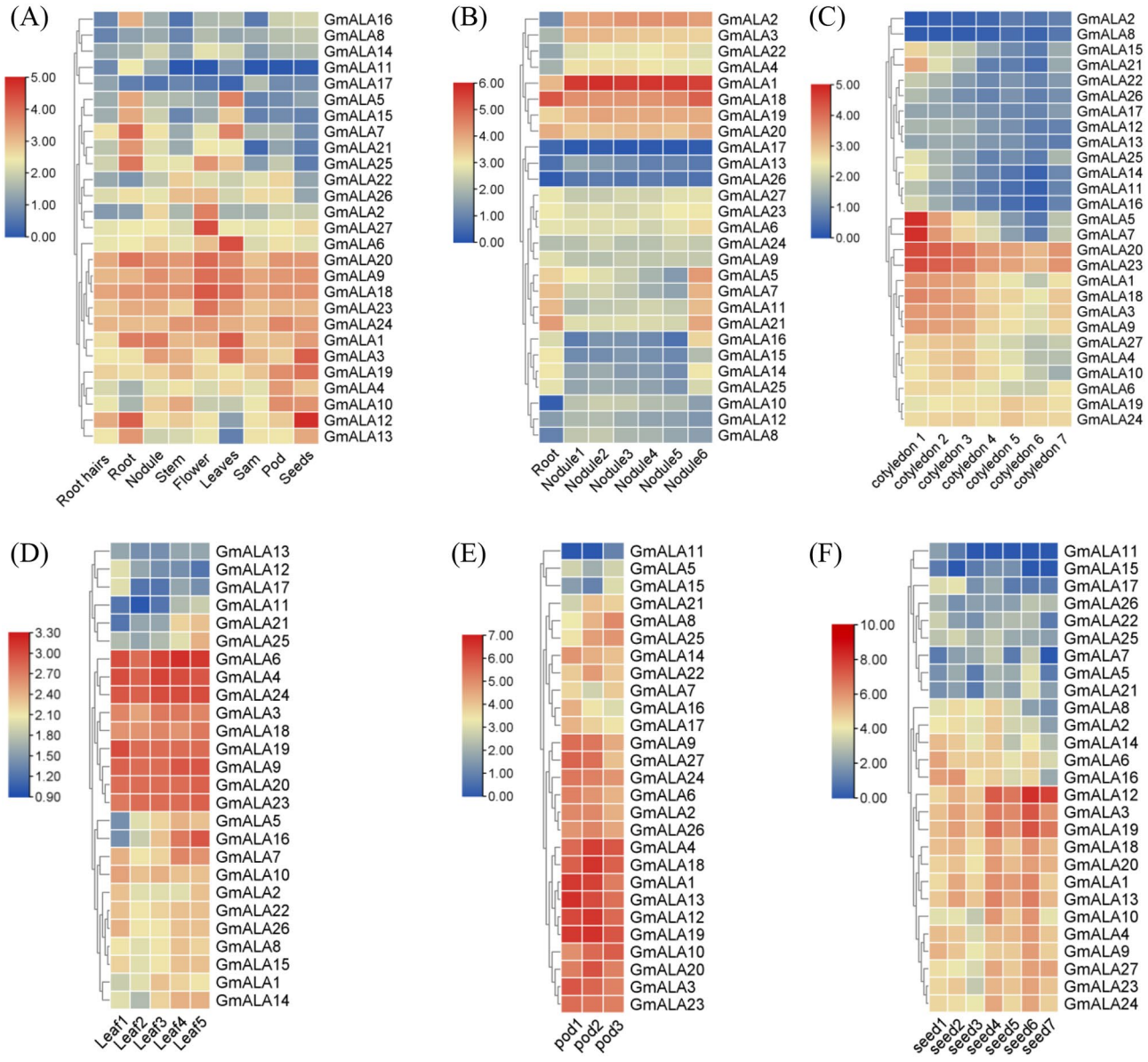


Fig. 5 Expression patterns of *GmALAs* in different tissues. (**A**) Tissue-specific expression of *GmALAs*. Different tissues (leaf, stem, root, root hair, nodule, sams, seeds, pods and flowers) were harvested from *G. max* cv. Williams 82 plants that had been grown for 4 weeks. The data were retrieved from phytozome database (https://phytozome-next.jgi.doe.gov/). (**B**) *GmALAs* expression in different developmental nodules. Nodules from the soybean cultivar 'Tianlong #.1' were harvested on different days after inoculation (10, 16, 22, 26, 32, and 38 days). The data were retrieved from the soybase (https://www.soybase.org/expression/). (**C**) *GmALAs* expression in cotyledons at different developmental stages. Seven different stages of cotyledons were collected during the development of soybean (*G. max* cv. Williams) seedlings. Stage 1, imbibed seed for 24 h; pre-emerging hypocotyls. Stage 2, yellow cotyledons; the emerging radicle is 8–10 mm long. Stage 3 included yellow cotyledons with slightly green edges and 15–20 mm long hypocotyls. In stage 4 yellow-green cotyledons and hypocotyls 30–35 mm long were observed. In stage 5, yellow-green cotyledons above the ground and primary roots start to develop. In stage 6, mostly green cotyledons above the ground grew straight from the hypocotyl. Stage 7, fully green cotyledons; plants 6–7 cm long above the ground; the root system fully developed; cotyledons upright; unifoliolate exposed. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**D**) *GmALA* expression in leaves at different developmental stages. Leaf samples were collected at different stages from Williams-82 soybean plants. Samples from stage 1 were collected after the trifoliates opened but before the leaves fully expanded. Stages 2 and 3 were collected 21 and 40 days after stage 1, respectively. Stages 4 and 5 were collected after 49 and 56 days, respectively. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**E**) *GmALA* expression in different developmental pods. The samples were collected from *G. soja* (PI468916) introgressed into *G. max* (A81-356022). Pod1: one cm pod, Pod2: pod shell 10 days after flowering, Pod3: pod shell 14 days after flowering. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**F**) *GmALA* expression in seeds at different developmental stages. Seeds (seed1, 10-DAF; seed2, 14-DAF; seed3, 21-DAF; seed4, 25-DAF; seed5, 28-DAF; seed6, 35-DAF; and seed7, 42-DAF) were collected on different days after flowering of the introgressed *G. soja* (PI468916) into *G. max* (A81-356022). The data were retrieved from the soybase (https://www.soybase.org/expression/)

GmALA27, *GmALA1*, *GmALA24* and *GmALA12* had the highest expression levels in leaves, roots, flowers, nodules, pods and seeds, respectively. Moreover, *GmALA18*, *GmALA20* and *GmALA12* also had higher expression levels than other *GmALAs* in root hair. In addition, *GmALA18*, *GmALA9* and *GmALA3* were higher expression levels than other *GmALAs* in sams. *GmALA9*, *GmALA18*, *GmALA1*, *GmALA19*, *GmALA3* and *GmALA7* were second highest expression than other *GmALAs* in the nodules, flowers, leaves, pods, seeds and roots, respectively.

GmALAs presented various expressed tendencies in different developmental tissues (Fig. 5B-F). Among them, *GmALA1* maintained relatively high expression in developmental nodules and most *GmALAs* exhibited staple expression except *GmALA5*, *GmALA7*, *GmALA11* and *GmALA21*, whose expression suddenly increased in senescent nodules (nodules 6) (Fig. 5B). Most of the *GmALAs* tended to be downregualted at different stages of cotyledon development, especially *GmALA5* and *GmALA7* whose expression decreased by approximately 90% in fully green cotyledons 1 compared with pre-emerging hypocotyls (cotyledons 7) (Fig. 5C). The remaining *GmALAs*, such as *GmALA2* and *GmALA24*, exhibited almost low or stable expression in developing cotyledons (Fig. 5C). Relative to those in developing cotyledons, the expression of one-third of *GmALAs*, such as *GmALA4* and *GmALA6*, remained high expression from opened trifoliates leaves to senescenced leaves (Fig. 5D). An increase or decrease in the expression of *GmALAs*, such as *GmALA16* and *GmALA12*, was also detected in developing leaves. The remaining *GmALAs* (*GmALA13*) exhibited relatively low expression and were stable at different leaf stage (Fig. 5D). Although the expression of most *GmALAs* was almost unchanged in the developmental pod, the expression of many *GmALAs*, such as *GmALA17* or *GmALA12*, increased or decreased in seeds of 10 to 35 days after flowering (Fig. 5E, F).

Response of *GmALAs* to biotic stress

GmALAs expression in leaves and roots after aphid infestation, fungal infection, or rhizobial treatment was obtained from public transcriptome data. The results demonstrated that all *GmALAs* were involved in these four biotic stresses and showed diverse expression patterns (Fig. 6). Further analysis revealed that some *GmALAs* had greater expression in the susceptible near-isogenic line (NIL) of Wyandot than in the resistant NIL after aphid infestation for 24 h, such as *GmALA1* and *GmALA3*, which both first increased but then decreased in expression (Fig. 6A). The other *GmALAs*, such as *GmALA11* and *GmALA15*, maintained low gene expression levels in the susceptible and resistant NILs and did not markedly differ between the aphid infected plants and the control plants (Fig. 6A). Both the nonpathogenic fungus *FO36* and the pathogenic fungus *FO40* induced

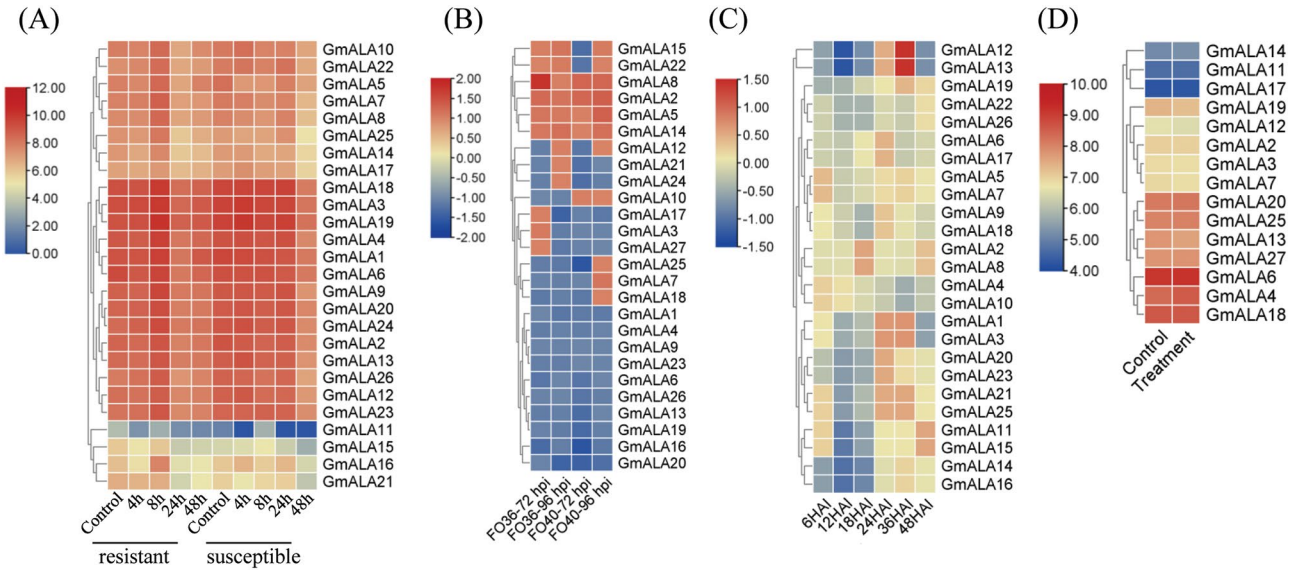


Fig. 6 Expression patterns of *GmALAs* in different tissues under biological stress. (**A**) *GmALA* expression in the leaves of Wyandot varieties near isogenic lines resistant and susceptible to aphid treatment. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**B**) *GmALA* expression in roots of plants treated with the non-pathogenic fungus (FO36) or pathogenic fungus (FO40) of *Fusarium oxysporum*. Hpi: hours post inoculation. Soybean [*G. max* (L.) Merrill] Forrest roots were collected at 72 and 96 h post inoculation. The data were retrieved from soyKB (https://soykb.org/). (**C**) *GmALA* expression in rhizobium treated roots. Soybean *G. max* (L.) Merr. cultivar 'Williams 82' plants roots were isolated at 6, 12, 18, 24, 36, and 48 h after incubation with *Bradyrhizobium japonicum* USDA110. The data were retrieved from soyKB (https://soykb.org/). (**D**) *GmALA* expression in soybean leaves treated with rust pathogen (*Phakopsora pachyrhizi Sydow*). The soybean [*Glycine max* (L.) Merrill] cv. 5601 T leaves were collected at 72 h post-inoculation. The data were retrieved from the soybase (https://www.soybase.org/expression/)

the expression of some *GmALAs*, such as *GmALA2*, *GmALA5*, and suppressed the transcriptional levels of some *GmALAs*, such as *GmALA1* and *GmALA4* (Fig. 6B). Interestingly, six *GmALAs*, *GmALA15*, *GmALA22*, *GmALA10*, *GmALA17*, *GmALA27*, and *GmALA3*, presented opposite expression patterns at 72 h in roots under *FO36* and *FO40* infection (Fig. 6B). In contrast to that in response to fungal infection, the *GmALAs* expression pattern in roots under rhizobial infection was more complex (Fig. 6C). Most *GmALAs*, except *GmALA12* and *GmALA13*, which both presented significant decreases at 12 h after incubation (HAI) and increases at 36 HAI, showed no obvious differences during the rhizobium infected period (Fig. 6C). Moreover, *GmALA14-16* and *GmALA11* also presented significant decreases at 12 HAI. It seemed that long term infection with Asian soybean rust (*Phakopsora pachyrhizi Sydow*) had no effect on the expression of *GmALAs* in leaves, and the expression patterns of some *GmALAs*, such as *GmALA17*, were similar to those in leaves under F40 infection (Fig. 6C, D).

Response of *GmALAs* to abiotic stress

Since abiotic stress-responsive elements were detected in the promoters of *GmALAs* (Fig. 4D), we also tested the expression of *GmALAs* in leaves or roots under dehydration, drought, blue light, wounding, saline, cold conditions and ozone exposure (Fig.S1). The expression profiles demonstrated that all *GmALAs* were differentially expressed under abiotic stress conditions (Fig. 7). However, the expression levels of most *GmALAs* in roots under natural dehydration showed little dramatic change, except for *GmALA11* and *GmALA13*, whose expression decreased and increased, respectively, at 1 h in dehydrated roots (Fig. 7A). Similar to natural dehydration treatment, all three genes (*GmALA9*, *GmALA18* and *GmALA23*) had extremely high and stable expression in

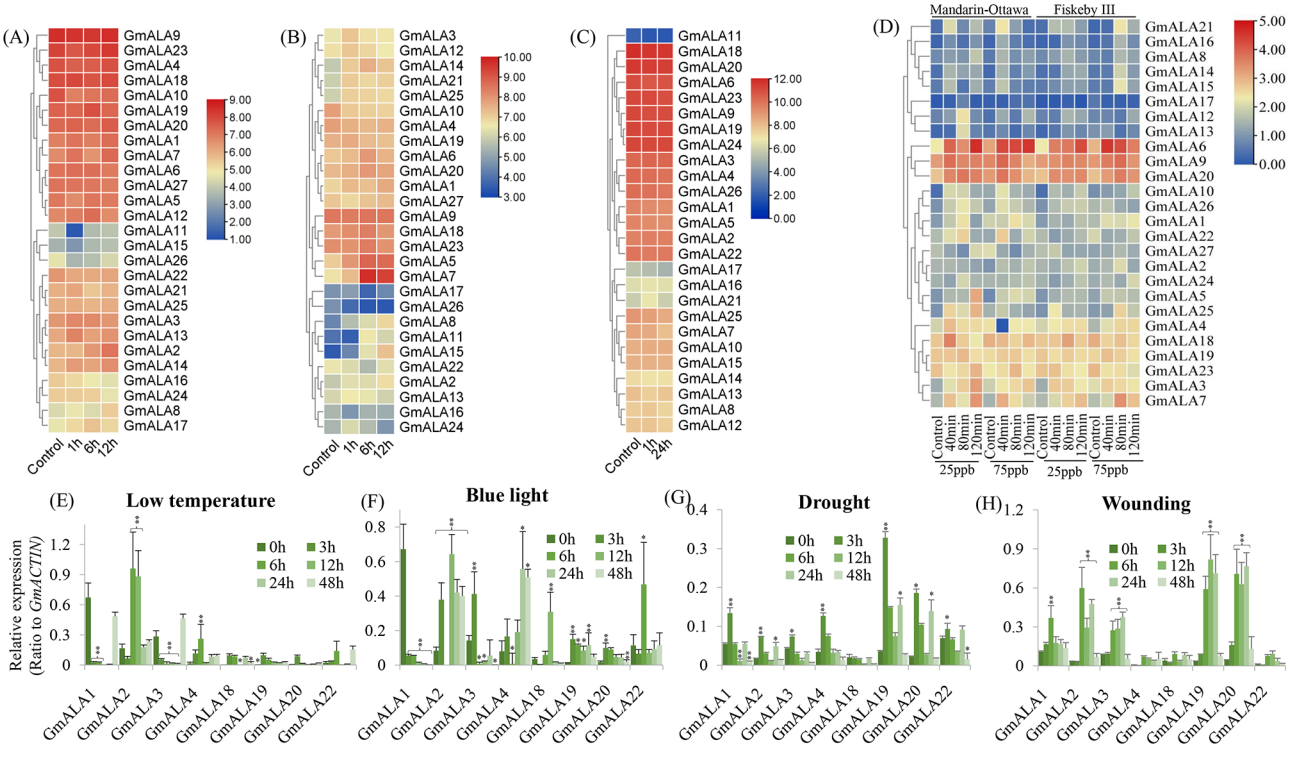


Fig. 7 Expression patterns of *GmALAs* in leaves and roots under abiotic stress. (**A**) *GmALA* expression in roots after different durations of dehydration stress. Root tissue was harvested after 0, 1, 6 and 12 h of *G. max* cv. Williams 82 plants were removed from the germination paper and left in air under water-limiting conditions to impose dehydration stress. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**B**) *GmALAs* expression in roots under different durations of salt stress. Root tissue was harvested after 0, 1, 6 and 12 h of *G. max* cv. Williams 82 plants were transferred to 100 mM NaCl solution. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**C**) *GmALAs* expression in leaves under different durations of low-temperature stress. Two-week-old Williams 82 unifolate seedling leaves were subjected to 4 °C for 1 or 24 h. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**D**) *GmALA* expression in the leaves of cultivars sensitive and tolerant to ozone. Sensitive Madarin-Ottawa and tolerant Fiskeby III cultivars were exposed to low (25 ppb) and high (75 ppb) ozone concentrations. The data were retrieved from the soybase (https://www.soybase.org/expression/). (**E**-**H**) Relative expression of *GmALAs* in leaves of Tianlong #1 plants under low temperature (4 °C), blue light, drought treatment (20% PEG6000) and injury treatment (two holes in each leaf). The results are presented as the means ± standard deviations. Samples were collected at 0, 3, 6, 12, 24 and 48 h and original samples were used as control. * and ** above the bars denote significant differences at $P < 0.05$ and $P < 0.01$, respectively, relative to the control

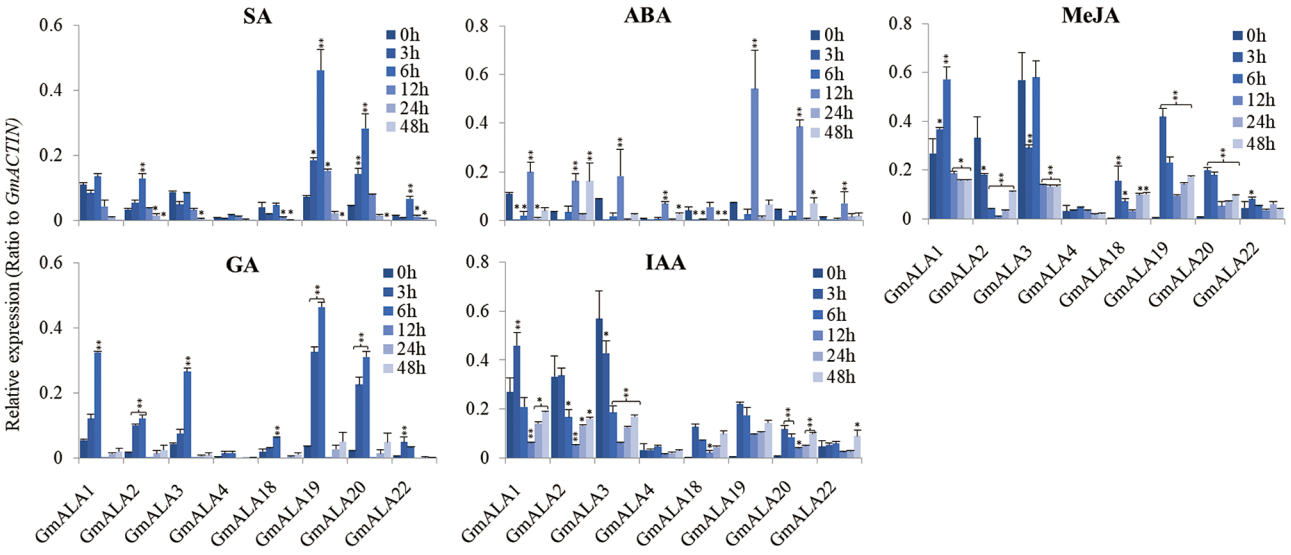


Fig. 8 Relative expression of *GmALAs* in leaves under five hormone stresses

roots under saline-stressed conditions (Fig. 7B). Moreover, all four *GmALAs* (*GmALA5*, *GmALA7*, *GmALA11* and *GmALA15*) were induced at 12 h and the remaining *GmALAs*, such as *GmALA2* and *GmALA13*, seemed to remain stable in roots during salt stress (Fig. 7B). Transcriptomes revealed that most *GmALAs* were consistent with the expression levels in leaves between the control and 4 ℃ treated groups at 1 h and 24 h (Fig. 7C). Moreover, detailed qRT-PCR analysis revealed that the expression of *GmALA2* and *GmALA4* increased by an average of 4-fold at 6 h compared with that of the control in the leaves under low temperature (4℃) (Fig. 7E). The three main *GmALAs* (*GmALA6*, *GmALA9* and *GmALA20)* had extremely high expression levels in leaves under ozone treatment, and they all increased slightly at 40 min compared with the control in the 25 ppb or 75 ppb ozone-intolerant cultivar (Mandarin-Ottawa) and the resistant cultivar (Fiskeby III) (Fig. 7D). Blue light, 20% PEG6000 treatment and physical damage caused different responses of the target *GmALAs* (Fig. 7F-H). For example, *GmALA1* tended to decrease during blue light, first increasing and then decreasing in leaves under drought and wounding stress (Fig. 7F-H). However, *GmALA2-4*, *GmALA19* and *GmALA20* showed similar patterns, with their transcriptional levels first increasing and then decreasing in leaves under these three conditions, in particular, *GmALA2* increased 7-fold at 12 h compared with that of the control in leave under blue light conditions. It seemed that *GmALA2* was promoted at 6 h in leaves during drought and wounding (Fig. 7G, H). In all, *GmALAs* were promoted or suppressed in different conditions, such as *GmALA1* was suppressed in 4℃ temperature or blue light condition at 6 h and promoted in drought or wounding condition at 3 h.

Response of *GmALAs* to phytohormones

Phytohormone-responsive elements were also detected in the promoters of *GmALAs* (Fig. 4D). To validate the function of *GmALAs* in response to hormone regulation, the expression of *GmALAs* in leaves under salicylic acid (SA), abscisic acid (ABA), auxin (IAA), methyl jasmonate (MeJA) and gibberellin (GA) treatments was detected (Fig.S2). The results showed that *GmALAs* exhibited various expression trends under phytohormone treatment (Fig. 8). The expression levels of targeted *GmALAs* (*GmALA1*, *3*, *4* and 18) were significantly decreased at 48 h in leaves under SA treatment. For example, the *GmALA1* level decreased by approximately 95% at 48 h compared with that of the control. All the remaining *GmALAs* (*GmALA2*, *19*, *20* and 22) increased at 12 h in leaves during SA treatment. In contrast to the SA treatment, all the selected *GmALAs* increased at 12 h–6 h in leaves under ABA or GA spraying respectively, for example, *GmALA19* increased approximately 7-fold in leaves at 6 h compared with control under GA treatment. The expression trends of all *GmALAs* in leaves under MeJA treatment were partially consistent with those under IAA treatment, for example, *GmALA2* was downregulated at 12 h under both conditions. Moreover, *GmALA19* increased approximately 9-fold and 4-fold at 3 h in response to MeJA and IAA, respectively.

The results are presented as the means ± standard deviations. Samples were collected at 0, 3, 6, 12, 24 and 48 h and the original samples were used as controls. * and ** above the bars denote significant differences at $P < 0.05$ and $P < 0.01$, respectively, relative to the control.

Discussion

P4-ATPase (ALA) gene family members are widely distributed in eukaryotes and have been primarily studied in mammals, yeast and a small number of plants [20–23]. There are 12 distinct P4-ATPases in *Arabidopsis thaliana*, while there are 8 in rice. Humans have 14 members of this family, while *S. cerevisiae* harbour 5 [5, 21–23]. In this study, we identified 27 *GmALA* genes that clustered into five main groups based on classification standards, which was greater than the number identified in *Arabidopsis*, rice and yeast [24, 25]. There was no direct correlation between the number of P4-ATPase genes and genome size in these plants, based on the sizes of the *Arabidopsis* (125 Mb), rice (430 Mb), and soybean (1.025 Gb) genomes. Gene duplication, determined through homology analysis, is considered to be one of the most important driving forces of genome evolution [26]. Generally, gene duplications include tandem repeats, segmental duplications and interspersed repeats, while segmental and tandem duplications are considered the main factors for gene family expansion in plants [27]. Tandem duplication and segmental duplications are important reasons for the generation and expansion of gene families in plants. Our research grouped five clusters from the ALA family of soybean, rice and *Arabidopsis thaliana*. This indicates that they may have arisen from a common ancestor. Syntenic analysis revealed that *GmALAs* showed pairwise synteny with genes in the *A. thaliana* and rice genomes, indicating that *GmALA*s underwent specific evolutionary events after the divergence of these three species. Most genes have undergone replication events, which increase the number of genes. This can be proven by the Ka/Ks value. A Ka/Ks value less than 1 demonstrates functional constraint due to purifying or negative selection of the genes [28]. We conclude that *GmALAs* experience strong purifying selection pressure with limited functional divergence after segmental duplications. Many orthologous gene pairs among the three species also suggest a high degree of similarity in gene sequences. Moreover, a total of 24 pairs of segmentally duplicated genes have been identified in the soybean genome, while zero pairs of tandemly duplicated genes have been identified, implying that segmental duplication events are the main source for the expansion of the *GmALA* family in soybean. This result is possibly due to the allotetraploid nature of soybean.

The evolution of gene families largely depends on the organization of gene structure. The varied length of the nucleotide sequence and distinct gene structures among *GmALAs* indicate the complexity of the soybean genome and divergent biological functions in soybean development. The molecular weights and isoelectric points of the GmALA proteins also differed among the family members, suggesting that their functions diverged. Additionally, GmALA proteins encompass 10 conserved motifs with varying compositions, and GmALAs members are grouped into the same subfamily based on similar motif types and counts, highlighting the balance between conservation and diversification within the GmALA family. The majority of *GmALAs* possess typical exons and introns, which exist in a splicing pattern and conserved PhoLip_ATPase_C, phosphorylation and transmembrane domains. This result is consistent with previous reports in other plants [15, 29]. All *GmALAs* are predicted to be expressed in the plasmolemma, and most GmALA proteins are neutral or weakly acidic, hydrophilic, and relatively stable. Based on the results above, *GmALAs* exhibit functional redundancy and divergence in the membrane during soybean development.

The potential *cis*-acting elements in the promoter region of *GmALAs* play a crucial role in initiating and regulating gene expression in various tissues under different environmental conditions. A total of 37 *cis*-elements were identified in the *GmALA* promoter region. Among these *cis*-elements, 7 are associated with cell development, 7 are involved in phytohormone responsiveness, 8 are related to light responsiveness, and 8 are associated with stress-related *cis*-regulatory elements. The function of the remaining *cis*-regulatory elements is unclear, indicating the diverse functions of the regulatory elements in *GmALAs*. These diverse *cis*-regulatory elements in the *GmALA* promoter region may also reflect functional differences at the transcriptional level.

ALAs play a crucial role in the transport of phospholipids and are involved in plant development and various stress responses [8, 15, 17]. For example, *AtALA6* is involved in the response to heat stress in *Arabidopsis thaliana* [24]. *Arabidopsis ALA1* and *ALA2* are responsible for mediating RNAi-based antiviral immunity, and *ALA1* also plays a role in chilling tolerance in *Arabidopsis* [14, 30]. *FgDnfA* and *FgDnfD* are crucial for the pathogenesis of *Fusarium graminearum*, and a significant decrease in deoxynivalenol (DON) production is observed in *ΔFgDNFA* and *ΔFgDNFD* [31]. In the present study, gene expression profiling data from different tissues (roots, nodules, stems, flowers, leaves, sams, pods and seeds), various abiotic stresses (drought, salt, ozone, low temperature, and physical damage), biotic stresses (aphid biting, fungus infection, rust and rhizobium treatment) and hormone treatments (salicylic acid, abscisic acid, IAA, MeJA and gibberellin) were used to dissect the functional roles of *GmALAs*. Different expression patterns are identified among *GmALA* genes, such as *GmALA1* is significantly expressed in different developmental nodules, which will be a candidate target for functional communication with nodule formation and development in soybean. These results indicate the functional differentiation of *GmALAs*. In addition, the expression of *GmALAs* in response

to hormone stress conditions was analyzed by qRT-PCR. The expression of a total of 4 *GmALAs (GmALA2, 19, 20 and 22)* could be significantly induced by SA, ABA and GA spraying at 6 h, 6 h and 12 h, respectively, suggesting that *GmALAs* may be involved in the cross-talk of different signaling pathways involved in hormone metabolism. Additionally, the expression difference of *GmALAs* indicates they possibly regulate the different lipids flipping under different environmental conditions and *GmALAs* family genes mutually coordinate the external signals to form charged lipids cavity to attract the intracellular protein or other target molecular in cytosolic side of biological membranes. Although genes with similar structures will be clustered in the same subfamily and may have similar biological functions, some *GmALAs* that are not clustered in the same group, such as *GmALA1* and *2*, show distinct response patterns under abiotic and biotic stress, suggesting that they may be involved in different response pathways under stress, and further experiments should be conducted to validate these functions. In summary, our results provide useful information for further functional exploration of these genes.

Conclusions

In this study, we characterized 27 *GmALAs* in the soybean genome, and these genes were divided into five groups. These genes had different motif compositions and gene structures. Moreover, all the *GmALAs* were distributed randomly on 15 chromosomes. A total of 24 pairs of *GmALAs* were identified from segmental duplications. *Cis*-regulatory elements of the *GmALA* promoters are involved in cellular development, phytohormones, environmental stress and photoresponsiveness. The *GmALA* family presented differential expression patterns in different tissues and developmental tissues, and differential responses were also found under different abiotic and biotic stresses for *GmALAs*. Interestingly, *GmALAs* are also involved in the regulation of hormone metabolism. Our results provide a valuable insight for improving the productivity and enhancing the environmental fitness of soybean plants.

Materials and methods

Identification of the *GmALAs* in the soybean genome

The *Arabidopsis thaliana* ALA protein sequence was used for homology alignment by TBlastP utilizing BLOSUM62 with an expect threshold = -1 and "# of alignments to show = 100 in the soybean proteome (version Wm82.a4.v1) from the database JGI Phytozome (http://phytozome.jgi.doe.gov/pz/ portal.html) to obtain sequences similar to those of members of the *AtALA* gene family [32]. The obtained candidate sequences with no conserved ALAase domain were deleted and gene family identification was performed using the SMART (http://smart.embl-heidelberg.de/) and Pfam (http://pfam.xfam.org/) databases. To further explore the characteristics of their domain-containing proteins, the ExPASY program (http://web.ExPASY.org/prot) was used to calculate the molecular weight (MW) and isoelectric point (pI), and the online software CELLO v.2.5 (http://cello.life.nctu.edu.tw/) was used to predict their subcellular localization [33].

Chromosomal location, phylogenetic analysis, gene structure, conserved motif/domain, *cis*-acting elements and collinearity analysis

Glycine max (Wm82) genome was downloaded from the JGI Phytozome database. The physical position and chromosomal distribution information of the *GmALAs* were obtained by using TBtools software (http://www.tbtools.com/) [34]. Phylogenetic relationships were constructed using amino acid sequences by MEGA7.0 with the neighbor-joining method (1000 bootstrap replications). The conserved motif, gene structure and conserved domain were determined with the TBtools based on the software protocol. Collinearity analysis of *ALAs* from soybean, *Arabidopsis thaliana* and rice was performed by TBtools using MCScanX plugins. The *cis*-regulatory elements in the promoter region (1500 bp upstream of the starting codon) of the *GmALAs* were searched by the online program PlantCARE (http://bioinformatics.psb.ugent.be/webtools/plantcare/html/) [35].

Retrieving transcriptome data of *GmALAs*

The transcriptome data of tissues and developmental tissues were retrieved from the JGI Phytozome (http://phytozome.jgi.doe.gov/pz/portal.html) or SoyBase databases (http://soybase.org/), respectively [32]. The fragments per kilobase of transcript sequence per millions base pairs sequenced (FPKM) values of the *GmALA* genes in nine tissues, different developmental tissues, biotic treatment and abiotic treatment were retrieved based on their ID number. A heatmap was generated using the heatmap function of the TBtools package [34]. Different tissues (leaf, stem, root, root hair, nodule, sams, seeds, pods and flowers) were harvested from *G. max* cv. Williams 82 plants that had been grown for 4 weeks. Developmental nodules from the soybean cultivar Tianlong #1 were harvested on different days after inoculation (10, 16, 22, 26, 32, and 38 days). Developmental cotyledons were collected during the development of soybean (*G. max cv*. Williams 82) seedlings. Developmental leaf samples were collected at different stages from Williams 82 soybean plants. Developmental pods and seeds were collected from *G. soja* (PI468916) introgressed into *G. max* (A81-356022). The expression data of *Bradyrhizobium japonicum* or *Fusarium oxysporum* treated roots of *G. max* (L.) cultivar 'Williams 82' or 'Forrest' were retrieved

from soyKB (https://soykb.org/). The expression data in leaves of Wyandot or 5601 T varieties treated by aphid or rust pathogen were retrieved from the SoyBase databases (http://soybase.org/). The expression data in Williams 82 leaves or roots under dehydration, salt, low-temperature stress were retrieved from the SoyBase databases (http://soybase.org/). The expression data in sensitive Madarin-Ottawa and tolerant Fiskeby III cultivars leaves or roots under dehydration, salt, low-temperature stress were retrieved from the SoyBase databases (http://soybase.org/). The expression data in leaves of sensitive Madarin-Ottawa and tolerant Fiskeby III cultivars for ozone stress were retrieved from the SoyBase databases (http://soybase.org/). All original data were uploaded in the Table S2 and taken the logarithm with a base of 2 for heatmap.

Calculation of Ka/Ks values

MCScanX was used for pairwise alignments of the paralogous nucleotide sequences. Ka (non-synonymous substitution rate) and Ks (synonymous substitution rate) were estimated using the simple Ka/Ks calculator in TBtools. The divergence time (T) was calculated using the formula: $T = Ks/2\lambda$, where the synonymous mutation rate λ was 6.161029 for soybean [36–38].

Plant materials and stress treatments

The soybean seeds of *Glycine max* cv. Tianlong #1 [gifted varieties developed by professor of Xinan Zhou working in Oil Crops Research Institute of the Chinese Academy of Agricultural Sciences, (OCRI)] were sterilized in 1% sodium hypochlorite for 1 min, washed three times in sterilized water, and then germinated in sterilized mixed soil (vermiculite: humus = 1:1) under a 14 h/10 h (light/dark) photoperiod, at 25 ℃, and 60% relative humidity. Uniformly grown fully unfolded plants with 4–5 leaves were subjected to stressed conditions, including low temperature (4 ℃), blue light (50–100 μmol m^{-2} s^{-1}), drought (20% PEG6000), physical damage, 2 mM salicylic acid,100 μM abscisic acid, 10 μM IAA, 20 μM MeJA and 50 μM gibberellin (Fig.S1, S2). Untreated plantlets were used as controls. Leaves for gene expression analysis were collected at 0, 3, 6, 12, 24 and 48 h after treatment, immediately frozen in liquid nitrogen, and then stored at – 80 ℃ prior to RNA extraction. Three biological replicates were obtained from each time point.

RNA isolation and RT-qPCR analysis

qRT-PCR analysis was conducted following a previously described method [39, 40]. Total RNA was isolated using a Total RNA Kit (TIANGEN, Beijing, China) according to the manufacturer's instructions. A total of 600 ng RNA was reverse transcribed using the HiFiScript cDNA Synthesis Kit (CWBIO, Jiangsu, China) according to the manufacturer's protocol. Following ten-fold dilution of the original cDNA, RT-qPCR was performed using Hieff® qPCR SYBR Green Master Mix (No Rox) (YEASEN, Shanghai, China) on a LightCycler 96 (Roche, Basel, Switzerland). All operational procedures were performed according to the manufacturer's instructions. The qRT-PCR were performed in a total volume of 20 μL, including 10 μL of SYBR Premix Ex Taq, 8.2 μL of ddH_2O, 1 μL of diluted cDNA, and 0.4 μL of each forward and reverse primer. The qPCR program was as follows: initiation with a 5 min denaturation period at 95 °C, followed by 40 cycles of 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 20 s. In total, 3 biological replicates and 2 technical replicates were performed for each reaction, and the relative expression of genes was calculated by the ΔCt method. All the primers used in the study are listed in Table S3.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11468-2.

Supplementary Material 1

Supplementary Material 2

Acknowledgements

We appreciate the patience of the reviewers and editors.

Author contributions

G.Z. and C.L. conceived and designed the experiments. J.W. conducted the main experiments and prepared the figures and tables. H.L., S.W. and X.L. contributed reagents/materials preparation. J.W. and G.Z. wrote and revised the manuscript. Y.Q. and Z.S. helped draft and polish the manuscript. All authors have approved the manuscript.

Funding

This research was financially supported by the National Natural Science Foundation of China (32101744) and the High Level Research Fund for Qualified People of Henan University of Technology (2021BS017).

Data availability

All data generated during this study are included in this published article and its supplementary information files. The DNA, protein sequence analysed during the current study are available in the NCBI (PV092034-092060).

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 20 March 2024 / Accepted: 10 March 2025
Published online: 20 March 2025

References

1. Villagrana R, López-Marqués RL. Plant P4-ATPase lipid flippases: how are they regulated? Biochim Biophys Acta Mol Cell Res. 2024;1871(1):119599.

2. Shin HW, Takatsu H. Phosphatidylserine exposure in living cells. Crit Rev Biochem Mol Biol. 2020;55(2):166–78.
3. Poulsen LR, López-Marqués RL, McDowell SC, Okkeri J, Licht D, Schulz A, et al. The *Arabidopsis* P4-ATPase *ALA3* localizes to the golgi and requires a beta-subunit to function in lipid translocation and secretory vesicle formation. Plant Cell. 2008;20(3):658–76.
4. López-Marqués RL. Lipid flippases in polarized growth. Curr Genet. 2021;67(2):255–62.
5. Andersen JP, Vestergaard AL, Mikkelsen SA, Mogensen LS, Chalat M, Molday RS. P4-ATPases as phospholipid flippases-structure, function, and enigmas. Front Physiol. 2016;7:275.
6. Davis JA, Pares RB, Bernstein T, McDowell SC, Brown E, Stubrich J, et al. The lipid flippases ALA4 and ALA5 play critical roles in cell expansion and plant growth. Plant Physiol. 2020;182(4):2111–25.
7. Sanz-Fernández M, Rodríguez-Serrano M, Sevilla-Perea A, Pena L, Mingorance MD, Sandalio LM, et al. Screening *Arabidopsis* mutants in genes useful for phytoremediation. J Hazard Mater. 2017;335:143–51.
8. Yang Y, Niu Y, Chen T, Zhang H, Zhang J, Qian D, et al. The phospholipid flippase *ALA3* regulates pollen tube growth and guidance in *Arabidopsis*. Plant Cell. 2022;34(10):3718–36.
9. McDowell SC, López-Marqués RL, Poulsen LR, Palmgren MG, Harper JF. Loss of the *Arabidopsis Thaliana* P_4-ATPase *ALA3* reduces adaptability to temperature stresses and impairs vegetative, pollen, and ovule development. PLoS ONE. 2013;8(5):62577.
10. López-Marqués RL, Davis JA, Harper JF, Palmgren M. Dynamic membranes: the multiple roles of P4 and P5 ATPases. Plant Physiol. 2021;185(3):619–31.
11. Guo Z, Lu J, Wang X, Zhan B, Li W, Ding SW. Lipid flippases promote antiviral Silencing and the biogenesis of viral and host SiRNAs in *Arabidopsis*. Proc Natl Acad Sci U S A. 2017;114(6):1377–82.
12. Poulsen LR, Lopez-Marques RL, Pedas PR, McDowell SC, Brown E, Kunze R, et al. A phospholipid uptake system in the model plant *Arabidopsis Thaliana*. Nat Commun. 2015;6:7649.
13. Vianna GR, Cunha NB, Rech EL. Soybean seed protein storage vacuoles for expression of Recombinant molecules. Curr Opin Plant Biol. 2023;71:102331.
14. Gomès E, Jakobsen MK, Axelsen KB, Geisler M, Palmgren MG. Chilling tolerance in *Arabidopsis* involves ala1, a member of a new family of putative aminophospholipid translocases. Plant Cell. 2000;12(12):2441–54.
15. Liu T, Guo S, Lian Z, Chen F, Yang Y, Chen T, et al. A P4-ATPase gene *GbPATP* of cotton confers chilling tolerance in plants. Plant Cell Physiol. 2015;56(3):549–57.
16. Wang F, Li X, Li Y, Han J, Chen Y, Zeng J, et al. *Arabidopsis* P4 ATPase-mediated cell detoxification confers resistance to *Fusarium graminearum* and *Verticillium dahliae*. Nat Commun. 2021;12(1):6426.
17. Ren H, Li X, Li Y, Li M, Sun J, Wang F, et al. Loss of function of VdDrs2, a P4-ATPase, impairs the toxin secretion and microsclerotia formation, and decreases the pathogenicity of *Verticillium dahliae*. Front Plant Sci. 2022;13:944364.
18. Fang C, Kong F, Soybean. Curr Biol. 2022;32(17):R902–4.
19. Liu S, Zhang M, Feng F, Tian Z. Toward a green revolution for soybean. Mol Plant. 2020;13(5):688–97.
20. Stanchev LD, Rizzo J, Peschel R, Pazurek LA, Bredegaard L, Veit S, et al. P-Type ATPase Apt1 of the fungal pathogen *Cryptococcus neoformans* is a lipid flippase of broad substrate specificity. J Fungi (Basel). 2021;7(10):843.
21. Pukyšová V, Sans Sánchez A, Rudolf J, Nodzyński T, Zwiewka M. *Arabidopsis* flippase *ALA3* is required for adjustment of early subcellular trafficking in plant response to osmotic stress. J Exp Bot. 2023;74(17):4959–77.
22. Dieudonné T, Kümmerer F, Laursen MJ, Stock C, Flygaard RK, Khalid S, et al. Activation and substrate specificity of the human P4-ATPase ATP8B1. Nat Commun. 2023;14(1):7492.
23. Pazos I, Puig-Tintó M, Betancur L, Cordero J, Jiménez-Menéndez N, Abella M, et al. The P4-ATPase Drs2 interacts with and stabilizes the multisubunit tethering complex TRAPPIII in yeast. EMBO Rep. 2023;24(5):e56134.
24. Niu Y, Qian D, Liu B, Ma J, Wan D, Wang X, et al. *ALA6*, a P(4)-type ATPase, is involved in heat stress responses in *Arabidopsis Thaliana*. Front Plant Sci. 2017;8:1732.
25. Sakuragi T, Nagata S. Regulation of phospholipid distribution in the lipid bilayer by flippases and scramblases. Nat Rev Mol Cell Biol. 2023;24(8):576–96.
26. Moore RC, Purugganan MD. The early stages of duplicate gene evolution. Proc Natl Acad Sci U S A. 2003;100(26):15682–7.
27. Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ. Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Syst Biol. 2005;54(3):441–54.
28. Yang J, Zhang B, Gu G, Yuan J, Shen S, Jin L, et al. Genome-wide identification and expression analysis of the R2R3-MYB gene family in tobacco (*Nicotiana tabacum* L). BMC Genomics. 2022;23(1):432.
29. Baxter I, Tchieu J, Sussman MR, Boutry M, Palmgren MG, Gribskov M, et al. Genomic comparison of P-type ATPase ion pumps in *Arabidopsis* and rice. Plant Physiol. 2003;132(2):618–28.
30. Zhu B, Gao H, Xu G, Wu D, Song S, Jiang H, et al. *Arabidopsis ALA1* and *ALA2* mediate RNAi-based antiviral immunity. Front Plant Sci. 2017;8:422.
31. Yun Y, Guo P, Zhang J, You H, Guo P, Deng H, et al. Flippases play specific but distinct roles in the development, pathogenicity, and secondary metabolism of *Fusarium graminearum*. Mol Plant Pathol. 2020;21(10):1307–21.
32. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, et al. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 2012;40(Database issue):D1178–1186.
33. Yu CS, Chen YC, Lu CH, Hwang JK. Prediction of protein subcellular localization. Proteins. 2006;64(3):643–51.
34. Chen C, Wu Y, Li J, Wang X, Zeng Z, Xu J, et al. TBtools-II: a one for all, all for one bioinformatics platform for biological big-data mining. Mol Plant. 2023;16(11):1733–42.
35. Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, et al. 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):325–7.
36. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. Science. 2000;290(5494):1151–5.
37. Lavin M, Herendeen PS, Wojciechowski MF. Evolutionary rates analysis of leguminosae implicates a rapid diversification of lineages during the tertiary. Syst Biol. 2005;54(4):575–94.
38. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2010;463(7278):178–83.
39. Zhang G, Yang J, Cui D, Zhao D, Li Y, Wan X, et al. Transcriptome and metabolic profiling unveiled roles of peroxidases in Theaflavin production in black tea processing and determination of tea processing suitability. J Agric Food Chem. 2020;68(11):3528–38.
40. Ma Q, Wang S, Tan H, Sun Z, Li C, Zhang G. Tissue-specific transcriptome analyses unveils candidate genes for flavonoid biosynthesis, regulation and transport in the medicinal plant ilex Asprella. Sci Rep. 2024;14(1):29999.

Publisher's note