



ARTICLE

Systematic Analysis of the *FLA* Gene Family and Expression Profiling in Soybean Varieties with Varying Stem Thickness

Mazin Ahmed Abdelraouf^{1,2}, Xiaoqi He¹, Hind Abdelmonim Elsanosi^{1,3}, Tiantian Zhu¹, Jinghui Shi¹, Ullah Habib¹ and Li Song^{1,*}

¹Joint International Research Laboratory of Agriculture and Agri-Product Safety, The Ministry of Education of China, Yangzhou University, Yangzhou, China

²Department of Agronomy, College of Agricultural Studies, Sudan University of Science and Technology, Khartoum, Sudan

³Faculty of Agriculture, University of Khartoum, Khartoum, Sudan

*Corresponding Author: Li Song. Email: songli@yzu.edu.cn

Received: 27 January 2026; Accepted: 09 May 2026; Published: 29 June 2026

ABSTRACT: The fasciclin-like arabinogalactan protein (FLA) family is involved in important plant wall formation and mechanical strength of the stems, and has never been systematically characterized in soybean (*Glycine max*), a huge crop in which stem lodging has been the cause of significant losses in yield. Here, we found that the soybean genome has 64 *GmFLA* genes, or a considerable increase over Arabidopsis, rice, and poplar, and these genes were grouped into three phylogenetic clusters (A, B, and C) that have varied domain structures. Evolutionary studies showed that duplication of segments was the most common cause of family expansion, with all the duplicated pairs of genes undergoing robust purifying selection. Tissue-specific expression profiling revealed that about a third of *GmFLA* genes is specifically expressed in stems. In nine stem-preferential genes assayed in six soybean varieties with varying stem thickness, the five stem-thick (diameter > 11 mm) varieties had a consistently lower expression of the five genes, *GmFLA62*, *GmFLA43*, *GmFLA50*, *GmFLA07*, and *GmFLA56*, than the thick-stem. These results are the first genomic resource of the *GmFLA* gene family in soybean and also determine certain candidate genes, which are different with the stem thickness, an important determinant of lodging. The findings of this study reveal a novel relationship between FLA gene expression and stem architecture, and suggest both as prospective targets for genetic improvement of lodging resistance and yield stability in soybean.

KEYWORDS: Soybean; FAS domain; AGP regions; GPI anchor signal; tissue-specific expression

1 Introduction

The yield of soybean is affected by a very large number of factors, yet one of the most common causes of yield loss in the entire world is lodging, which refers to the permanent loss of stems in an upright position [1–3]. Understanding the molecular processes that regulate stem growth and mechanical strength is essential for developing lodging-resistant soybean varieties [4]. Given the importance of soybean to global agriculture, improving its productivity could have a transformative impact, promoting sustainable farming by enabling higher yields on existing land. The rise in demand for sustainable agriculture has led to interest in new, sustainable ways to increase production in the face of climate change and environmental pressures [5]. Plant cell walls are dynamic and complex structures that exhibit remarkable plasticity in response to environmental, mechanical, and biological stresses [6]. The cell wall is not merely a static barrier, but a dynamic structure that actively adapts to changing conditions to ensure the

cell's survival and function, a capability crucial for plant survival [7,8]. It also modulates plant defense signals under stress [9,10]. This dynamic nature stems from cell wall's composition, which includes cellulose, hemicelluloses, pectin polysaccharides, and about 10% glycoproteins abundant in proline and/or hydroxyproline residues. Hydroxyproline-rich glycoproteins (HRGPs) are defined by their abundance of proline and/or hydroxyproline residues. This class comprises diverse proteins such as extensins, arabinogalactan proteins (AGPs), glycine-rich proteins (GRPs), solanaceous lectins, and glycosylated proline-rich proteins (PRPs) [11]. AGPs represent the most complex and diverse subfamily of HRGPs and are found throughout the plant kingdom [12–14]. Through integrated approach, *AGP* gene families have been identified in 47 plant species, paving the way for a deeper understanding of AGPs physiological function [15]. The provided investigation suggests that AGPs play a crucial performance in plant development, especially in maintaining cell wall structure and signaling processing. AGPs are crucial for multiple cell wall functions, including cell expansion and division, signaling, cell-cell recognition and adhesion, cellulose synthesis and deposition, and modulation of cell wall mechanics [16,17].

Fasciclin-like AGPs (FLA) belong to the AGP subfamily. Most identified FLAs possess an N-terminal signal peptide, a C-terminal glycosyl phosphatidylinositol (GPI) membrane anchor, or both [11,13]. Bioinformatics methods have been used to identify and predict *FLA* gene family in various plant genomes, including *Arabidopsis thaliana* [18], rice (*Oryza sativa*) [19,20], cotton (*Gossypium hirsutum*) [21], Chinese cabbage (*Brassica rapa*) [22], eucalyptus (*Eucalyptus grandis*) [23,24], poplar (*Populus trichocarpa*) [25], and textile hemp (*Cannabis sativa*) [26]. The genome wide research demonstrated that *FLA* genes are highly identified among different crops species, however, their functional diversification may be species specific, especially in related to cell wall growth and physiological characteristics.

Functional studies have demonstrated that FLAs play critical roles in stem development and cell wall architecture, with direct implications for mechanical strength and, consequently, yield potential [23–25,27]. For instance, *AtFLA4*(*SOS5*) caused abnormal cell growth, adhesion, cell wall formation, and seed coat pectin mucilage in *Arabidopsis* [27]. *AtFLA11* and *AtFLA12* influence stem tensile strength, biomechanics, and elastic modulus, thereby affecting cell wall composition and structure [28]. Deletion of *AtFLA16* in *Arabidopsis* mutants resulted in reduced stem length and altered biomechanical properties [29]. Studies on poplar showed that the *PtFLA6* gene was associated with the poplar's xylem fiber cells, cell wall composition, and stem biomechanics, and *PtFLAs* have been implicated in gibberellin-mediated tension wood formation [30]. In eucalyptus, *EgFLA* genes affect wood properties and stem stiffness [23,24]. Collectively, these findings indicate that FLAs play critical roles in stem development and cell wall architecture that contribute to lodging resistance and, by extension, yield stability. Nevertheless, most of these investigation have emphasized on model plants or woody plants varieties, with limited research in major crops species such as soybean.

Despite these advances, several critical knowledge gaps remain. First, no systematic characterization of the *FLA* gene family has been performed in soybean (*Glycine max*), a crop where stem lodging is a major yield constraint. Previous research demonstrated that genome-wide *FLA* analyses have been conducted in various species; however, their performance in soybean is still in an early stage. Second, while *FLAs* have been linked to stem biomechanics in model and woody species, it is unknown whether soybean *FLAs* exhibit similar functions, and whether their expression correlates with stem thickness—a trait directly associated with lodging resistance and yield potential. Third, no upstream regulators of *FLA* genes have been experimentally confirmed, and co-expression networks remain unexplored in soybean.

We therefore hypothesize that the soybean *FLA* gene family has been evolutionarily expanded and subfunctionalized, and that specific *GmFLA* genes contribute to stem growth, mechanical strength, and

ultimately lodging resistance. In particular, we propose that differentiation in *GmFLA* gene expression is linked with stem thickness, showing a potential molecular link to lodging resistance. Specifically, we propose that (i) *GmFLA* genes exhibit tissue-specific and developmentally regulated expression patterns, and (ii) variation in their expression levels correlates with stem thickness, a critical determinant of lodging resistance and yield potential in soybean.

To test this hypothesis, this study aims to: (1) systematically identify and characterize the *GmFLA* gene family across the soybean genome using phylogenetic, structural, and duplication analyses; (2) profile the expression patterns of *GmFLA* genes across multiple tissues, with a particular focus on stem development; and (3) identify candidate *GmFLA* genes whose expression is correlated with stem thickness, thereby providing a foundation for future functional validation. The present study showed the novelty of associated genome-wide FLA analysis with stem thickness attributes in soybean plants. By establishing this genetic resource, our work bridges the gap between functional studies of FLAs in other species and the yet-unexplored potential of soybean FLAs to improve stem architecture and lodging resistance.

2 Materials and Methods

We thoroughly characterized the *GmFLA* gene family in soybean using genome-wide identification, phylogenetic analysis, duplication analysis, expression profiling and qPCR confirmation. The detailed methods are as follows:

2.1 Identification of FLAs Genes in the Soybean Genome

To identify members of the *FLA* gene family in soybean, protein sequences of *Arabidopsis FLA* genes were downloaded from Phytozome (<https://phytozome.jgi.doe.gov/pz/portal.html>, v13.1). These sequences, representing distinct FLA subfamilies, were used as queries in BLASTP searches (<http://blast.ncbi.nlm.nih.gov>) against the *Glycine max* reference genome (Wm82.a6.v1 version) applying an E-value threshold of $1e-6$. The search parameters included a gap opening penalty of 11, a gap extension penalty of 1, the BLOSUM62 substitution matrix, a word size of 6, and low-complexity region filtering enabled. Furthermore, a keyword search using “*GmFLA*” was performed in the Phytozome database against the same soybean reference genome. The resulting sequences were subjected to domain validation using Pfam databases (<http://pfam.jouy.inra.fr>) with an E-value cutoff of 1.0 to confirm the presence of the fasciclin domain (Pfam accession: PF02469). Domain verification was further carried out using the SMART databases (<http://smart.embl-heidelberg.de>) with an E-value cutoff level of 1.0. Any candidate sequence lacking this essential domain were discarded. For each confirmed *GmFLA* protein, the number of amino acids (aa), theoretical isoelectric point (pI), and molecular weight (MW) were calculated using the ExPASy online server (<https://web.expasy.org/protparam>). The Plant-mPLOC server (<http://www.csbio.sjtu.edu.cn/cgi-bin/PlantmPLOC.cgi>) was used to predict the subcellular localization. Plant-mPLOC combines information about functional domains and phylogenetic profiles to make predictions about subcellular targeting, and the total reported accuracy is around 80% with regard to plant proteins.

A protein sequence was defined as containing an AGP-like glycosylated region if it possessed two or more contiguous [A/S/T] P, [A/S/T] PP, and [A/S/T] PPP motifs, excluding the fasciclin domains and N- or C-terminal signals. SignalP5.0 was employed to predict the N-terminal GPI anchor signal peptide (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>), while the C-terminal GPI anchor addition signal was predicted using the big-PI Plant Predictor (http://mendel.imp.ac.at/gpi/plant_server.html) [20]. Conserved regions (H1, H2), together with motifs and residues relevant to adhesion, were identified manually.

2.2 Sequence Alignment and Phylogenetic Analysis

For classification purposes, the protein sequences of 64 *GmFLA* genes were retrieved from Phytozome v13.1. These sequences were then aligned using MUSCLE algorithm implemented in MEGA 11 [31]. An unrooted phylogenetic tree was generated with MEGA 11 employing the neighbor-joining (NJ) method, based on the p-distance approach with 1000 bootstrap replications. The resulting tree was further visualized and refined using the iTOL web platform (<https://itol.embl.de>) [32]. Sequence alignment was performed and examined using Jalview software [33].

2.3 Chromosomal Localization and Ka/Ks Ratios Analysis

The Chromosomal locations of *GmFLA* genes were visualized using TBtools [34] based on their physical mapping coordinates. Gene duplication events were analyzed with the Multiple Collinearity Scan toolkit (MCScanX) integrated within TBtools. Circos diagrams were generated with TBtools to illustrate duplication events and syntenic relationships among genes. Genes located on unassembled genomic scaffolds were omitted from further analyses. The ratio of nonsynonymous-to-synonymous substitution rates (Ka/Ks) was calculated using TBtools to evaluate the selective pressures acting on duplicated gene. A Ka/Ks value equal to 1 indicates neutral evolution, a value less than 1 signifies purifying selection, and a value greater than 1 suggests positive selection.

2.4 Gene Structure and Conserved Motif Analysis

The structure organization of *GmFLA* gene was analyzed using TBtools software [34]. Conserved motifs within *GmFLA* gene sequences were analyzed via MEME online tool (<http://meme-suite.org/tools/meme>, v5.3.3), The MEME analysis was done using the following parameters; classic (normal) discovery mode, maximum number of motifs to be discovered is 20, minimum motif width is 6 amino acids, maximum motif width is 50 amino acids, and E-value is less than 0.05, which is considered to be statistically significant. A Markov background model of 0-order was used. Motifs that had E-value more than 0.05 were not analyzed further. The identified motifs were plotted in *GmFLA* proteins using the TBtools software [35]. A diagram illustrating the motif distribution of these motifs was generated using TBtools, with motif pattern information derived from the XML output file produced by MEME [36].

2.5 Cis-Acting Elements Analyses

Promoter sequences of *GmFLA* genes were extracted from the soybean reference genome (Glycine max Wm82.a6.v1) using TBtools software [34]. For each gene, a 1500 base pairs (bp) region immediately upstream of the translation start codon (ATG) was extracted. Cis-regulatory elements present in these promoter regions were predicted using the PlantCARE database (<http://bioinformatics.psb.ugent.be/webtools/plantcare/html/>) and were subsequently manually annotated and categorized according to their putative biological functions. Each unique cis-element was recorded only once per promoter to avoid overcounting due to identical or overlapping sequences.

2.6 Expression Profiles of *GmFLA* Genes in Soybean Tissues

Expression patterns of *GmFLA* genes were examined using FPKM data retrieved from Phytozome, which included seven distinct tissues (root, lateral root, stem, leaf, shoot tip, opened flower and unopened flower). Heatmaps were generated with TBtools' heatmap plotting functionality to visualize the expression levels across these tissues.

2.7 qPCR Analysis

To compare gene expression patterns between stem thickness groups, three thick-stemmed varieties [JSL017 (11.46 ± 2.62 mm), JSC041 (11.68 ± 2.20 mm), JSL079 (11.58 ± 2.41 mm)] and thin-stemmed varieties [JSC028 (6.19 ± 1.15), JSL094 (4.69 ± 0.49 mm), JSL108 (5.71 ± 1.09 mm)] were used [37]. Seeds were planted on dishes with 1/2 MS medium for 10 days and hypocotyl tissues were then collected. Total RNA was isolated using RNeasy Plant Kit (Cat#CW0588S, CWBIO, Beijing, China), following the manufacturer's instructions. cDNA was synthesized using the HiScript 1st Strand cDNA Synthesis Kit (Vazyme, Cat#R111-01, Nanjing, China). Quantitative RT-PCR was performed using the Bio-Rad CFX Connect™ Optics Module Real-Time PCR System (Bio-Rad, CA, USA) and SuperStar Universal SYBR Master Mix (CWBIO, Cat#CW3360H, Taizhou, China). The qPCR reactions were performed with the following cycling parameters: initial denaturation at 95°C for 15 min; 40 cycles of denaturation at 95°C for 15 s, annealing and extension at 60°C for 30 s; followed by a melting curve analysis (95°C for 15 s, 60°C for 1 min, and 95°C for 15 s) to confirm the specificity of the amplified products. Reference genes *GmACTIN11* (Glyma.18G290800) were used for normalization. All primers used in this study (listed in Table S1) were synthesized by [Sangon Biotech (Shanghai) Co., Ltd.]. Relative expression levels were calculated using the $2^{-\Delta\Delta C_t}$ method. Nine plants for each variety (9 plants \times 6 variety) in three biological replicates were analyzed for each variety. An unpaired two-tailed Student's *t*-test was used to compare expression levels between the two groups, with $p < 0.05$ considered statistically significant.

3 Results

In total, we identified and defined 64 *GmFLA* genes. Below is provided the evolutionary classification, molecular characteristics, chromosomal distribution, duplication events, conserved motifs, cis-regulatory elements, tissues-specific expression patterns and qPCR based expression analysis for stem thickness.

3.1 Phylogenetic Classification of *GmFLA* Proteins

The classification of *GmFLA* proteins was based on phylogenetic analysis, resulting in the division of 64 *GmFLAs* into three distinct groups (Group A, B, and C; Fig. 1 and Table S2). Group A represents the largest group. Most members of group A contain a single fasciclin domain, a GPI-anchored signal, and more than two AGP regions. Exceptions include *GmFLA35*, *GmFLA39*, *GmFLA28*, *GmFLA25*, and *GmFLA01*, which possess only a single fasciclin domain and lack a GPI-anchor signal; *GmFLA53*, *GmFLA03*, *GmFLA52*, *GmFLA16*, *GmFLA60*, *GmFLA19*, and *GmFLA61*, contain two fasciclin domains, GPI-anchored signals, and multiple AGP regions; and *GmFLA01* possess a single fasciclin domain but lacks both a GPI anchor and an AGP region.

Group B exhibits heterogeneous domain organization, with variable absence or presence of a GPI anchor and AGP regions. Among this group, five proteins (*GmFLA21*, *GmFLA64*, *GmFLA18*, *GmFLA14*, *GmFLA09*, and *GmFLA48*) contain two fasciclin domains and lack a GPI-anchored signal and more than two AGP regions. Two proteins (*GmFLA15* and *GmFLA54*) contain a single fasciclin domain and lack both a GPI-anchored signal and more than two AGP regions. One protein (*GmFLA59*) contains single fasciclin domains, a GPI-anchored signal, and two AGP regions.

Group C corresponds to the smallest group. Most members contain two fasciclin domains, a GPI-anchored signal, and more than two AGP regions. The only exception is *GmFLA55*, which contains a single fasciclin domain and more than two AGP regions but lacks a GPI-anchor signal.

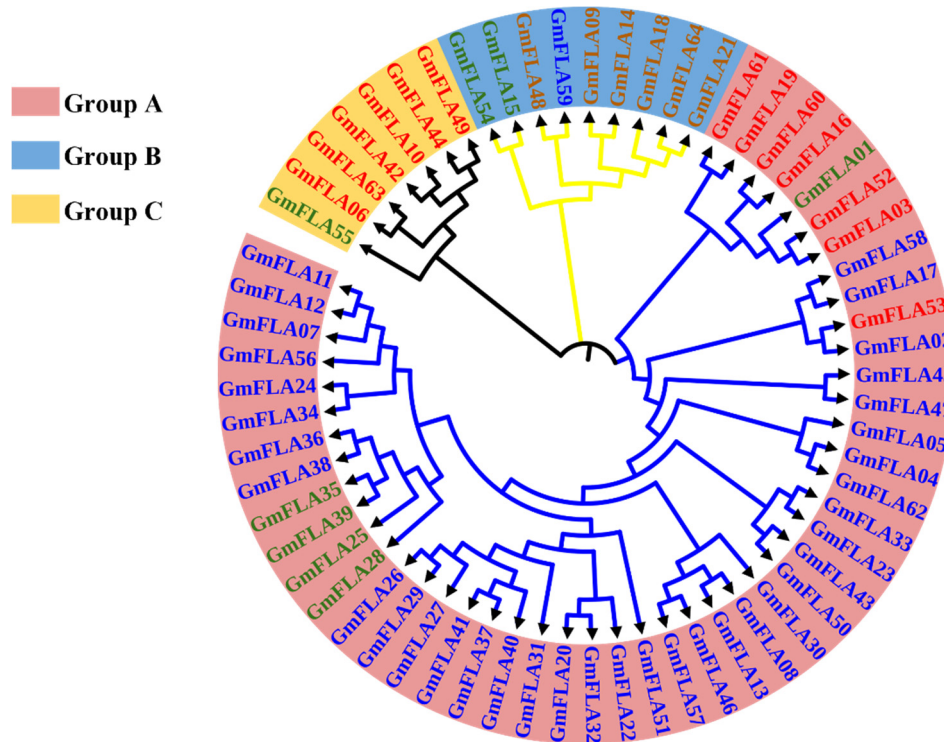


Figure 1: An unrooted phylogenetic tree of *GmFLA*. The deduced full-length amino acid sequences were utilized to construct the phylogenetic tree using MEGA 11 software through a neighbor-joining method with 1000 bootstrap replicates. Various groups are distinguished by colors: Group (A) in pink, Group (B) in blue, and Group (C) in yellow. The color of each gene name indicates its domain composition and the presence or absence of a GPI anchor: blue (single domain with C-terminal GPI anchor), green (single domain without C-terminal GPI anchor), red (two domains with C-terminal GPI anchor), brown (two domains without C-terminal GPI anchor).

3.2 Molecular Characteristics of *GmFLA* Genes

The attributes of the 64 *GmFLA* genes, including protein length, pI, MW, GRAVY and predictive subcellular localization, are listed in Table S3. The lengths of *GmFLA* proteins varied from 171 (*GmFLA01*) to 488 (*GmFLA01*) amino acids, with group C *GmFLA* members generally being longer than those in other groups. The theoretical pI values ranged from 4.25 (*GmFLA21*) to 9.56 (*GmFLA50*). The MW of *GmFLA* proteins ranged from 19.12 to 75.57 kDa. Regarding hydrophobicity, 16 *GmFLA* proteins exhibited negative GRAVY values, indicating hydrophilic behavior, while the remaining *GmFLA* proteins exhibited hydrophobic characteristics due to their positive GRAVY values. The subcellular localization predictions suggested that 44 *GmFLA* proteins are likely located to the extracellular space, nine are predicted to be nuclear, ten are chloroplast-localized, and one protein (*GmFLA44*) is present in the cytoplasm (Table S3). Experimental validation is required for confirm these predicated subcellular localizations.

3.3 Chromosomal Position, Gene Duplication and K_a/K_s Evaluate

The 64 *GmFLA* genes are unevenly distributed across the soybean chromosomes. No *FLA* gene were detected on chromosomes 1, 4, 7, 16 and 17. In contrast, chromosomes 11 and 12 harbor relatively large numbers of genes, ranging from 10 to 14, respectively, while chromosomes 20 and 6 have only one gene each (*GmFLA64* and *GmFLA10*, respectively; Fig. 2).

Small hydrophobic amino acids such as leucine, valine, and isoleucine are abundant in the H2 region. However, the H2 domain was not clearly identifiable in *GmFLA01*, *GmFLA09*, *GmFLA14*, *GmFLA15*, *GmFLA18*, *GmFLA21*, *GmFLA49*, *GmFLA54*, and *GmFLA55*. Within the [Y/F] H motif, histidine and proline residues are also relatively conserved. Only five *GmFLA* proteins lack [Y/F] H motif, namely *GmFLA09* (EH), *GmFLA14* (EH), *GmFLA18* (RH), *GmFLA54* (SH), and *GmFLA55* (IV).

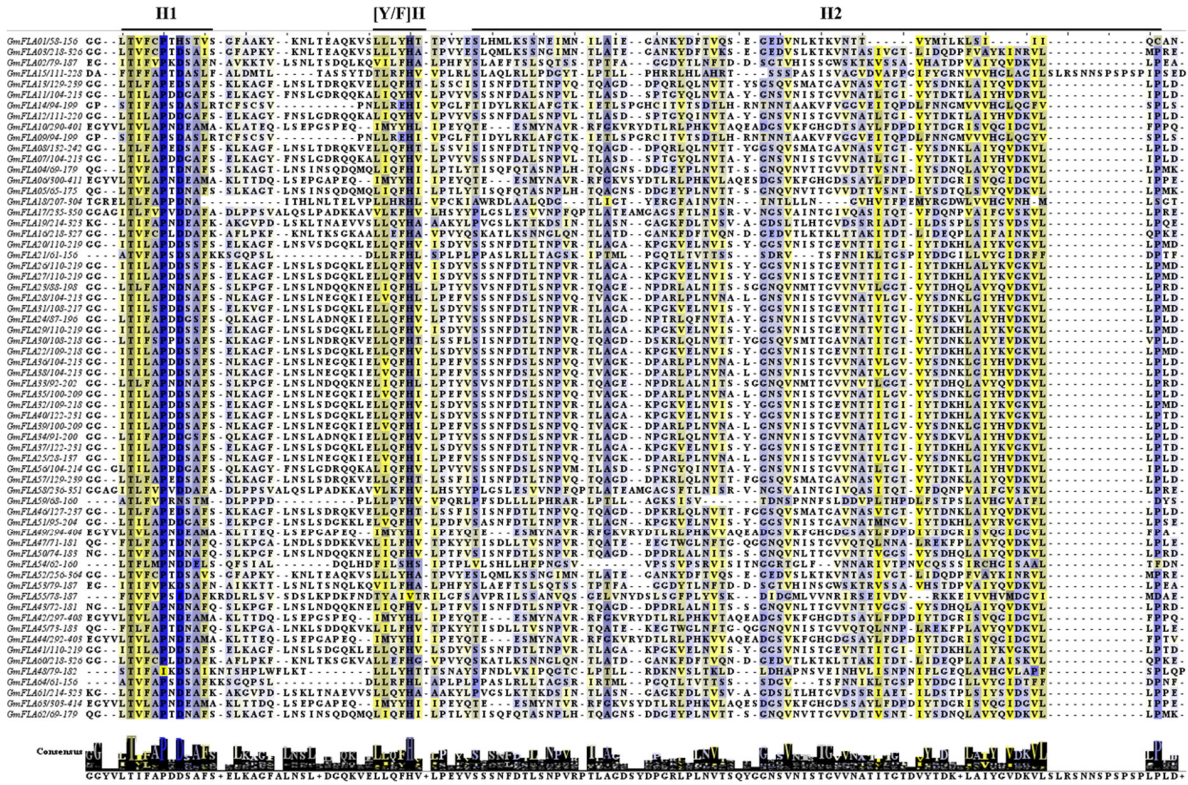


Figure 3: Multiple sequence alignment of the fasciclin (FAS) domain from 64 *GmFLA* proteins. Sequence alignment was performed using MUSCLE and visualized with Jalview. Conserved regions are highlighted: H1, H2 and [Y/F]H motif.

3.5 Gene Structure and Conserved Motifs Analysis of *GmFLA* Genes

The gene structure and conserved motif of the *GmFLA* gene family were further analyzed (Fig. 4). Analysis of genomic DNA sequences revealed that *GmFLA* genes typically contain zero or one intron. All members of group A lack introns, except for *GmFLA58* and *GmFLA17*, each of which possesses one intron. Similarly, all members of Group B lack introns, except *GmFLA14*, *GmFLA15* and *GmFLA09*. In group C, almost all members contain one intron, with the exception of *GmFLA10* and *GmFLA55*, which lack introns.

Twenty conserved motifs were identified across the *GmFLA* family. Each *GmFLA* protein contains between two and nine conserved motifs. Note that most of the members in group A contains motifs 3, 5, 2, 16 and 11. In addition, motif 2 is unique to group A. In contrast, motifs 13 and 4 are exclusive to group C. Motif 14 and 5 can be found in both group A and B.

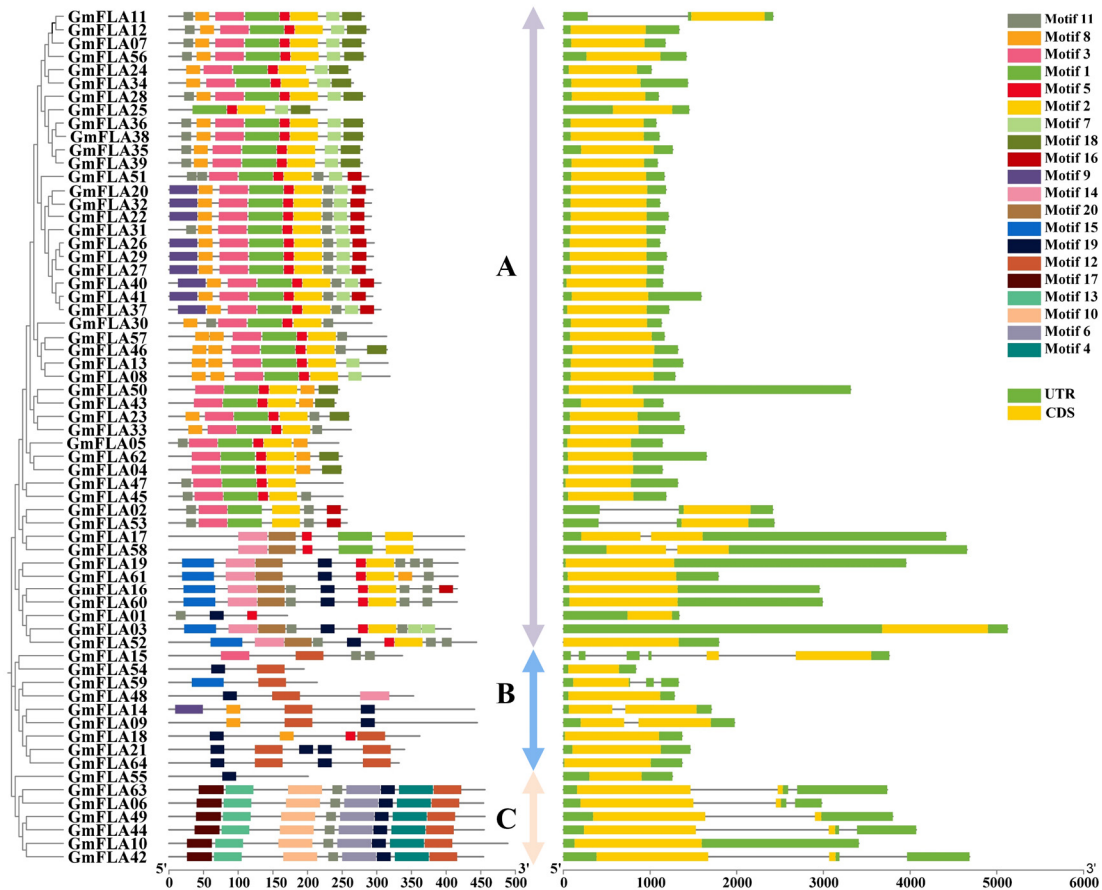


Figure 4: Phylogenetic relationship, conserved motif architectures, and exon–intron structure of *GmFLA* genes. Left: Phylogenetic tree showing the three major groups (A, B, C). Middle: Distributions of conserved protein motifs. Right: Exon–intron structures, with yellow boxes representing coding exons and black lines representing introns.

3.6 Cis-Acting Elements Analysis

Cis-regulatory elements are present in the upstream of transcriptional start sites, regulating the expression of the genes. A total of 137 cis-elements were discovered in the promoter regions, which were categorized into seven main functional groups, including light-responsive, MeJA-responsive, gibberellin-responsive, abscisic acid responsive, salicylic acid-responsive, environment-responsive and auxin-responsive (Fig. 5). The promoters of the *GmFLA* gene family are enriched in elements responsive to light and plant hormones (such as MeJA, ABA, gibberellin, auxin, and salicylic acid), as well as stress-related elements (Table S6). This enrichment suggests that *GmFLA* gene expression is tightly regulated by complex environmental and hormonal signals rather than being constitutive. The presence of MeJA- and ABA-responsive elements to a pivotal role for these genes at the intersection of stress response pathways and developmental processes—a finding consistent with the emerging evidence that cell wall proteins coordinate growth signals with defense mechanisms. Furthermore, the enrichment in elements responsive to gibberellin and auxin holds particular significance, given that these act as key regulators of stem elongation and secondary cell wall formation. This observation supports the hypothesis that *GmFLA* genes contribute to structural development of the stem. Additionally, the abundance of light-responsive elements aligns with the anticipated functions of these genes within photosynthetic and supporting tissues.



Figure 5: Prediction of cis-acting elements in the promoters of *GmFLA* genes. A number of cis-acting elements were divided into seven distinct functional categories: (1) MeJA-responsive (dark green), (2) Light-responsive (yellow), (3) Gibberellin-responsive (pink), (4) Abscisic acid-responsive (cyan blue), (5) Salicylic acid-responsive (red), (6) Environment-responsive (purple), and (7) Auxin-responsive (light green).

3.7 Expression Patterns of *GmFLA* Genes in Soybean Tissues

Expression analysis of the 64 *GmFLA* genes was conducted across seven soybean tissues: open flower, unopen flower, lateral root, root, shoot tip, stem, leaf (Fig. 6). The analysis revealed diverse, tissue-specific expression. Seven genes (*GmFLA01*, *GmFLA03*, *GmFLA17*, *GmFLA33*, *GmFLA42*, *GmFLA52*, and *GmFLA54*) exhibited no detectable expression (FPKM = 0) in any of the tissues examined, suggesting they may represent pseudogenes or have highly specialized expression under specific conditions or below detection thresholds.

Among the expressed genes, several showed tissue-preferential expressions. For example, *GmFLA07* and *GmFLA56* were highly expressed in stem (FPKM = 70.69 and 7.99, respectively) and root (FPKM = 33.04 and 1.39, respectively) compared with other tissues. Notably, *GmFLA46* and *GmFLA48* exhibited markedly high expression specifically in leaf, with FPKM values of 9.99 and 0.29, respectively, representing the highest expression levels for these genes among all tissues examined. A complete list of FPKM values for all genes across all tissues is provided in Table S7.

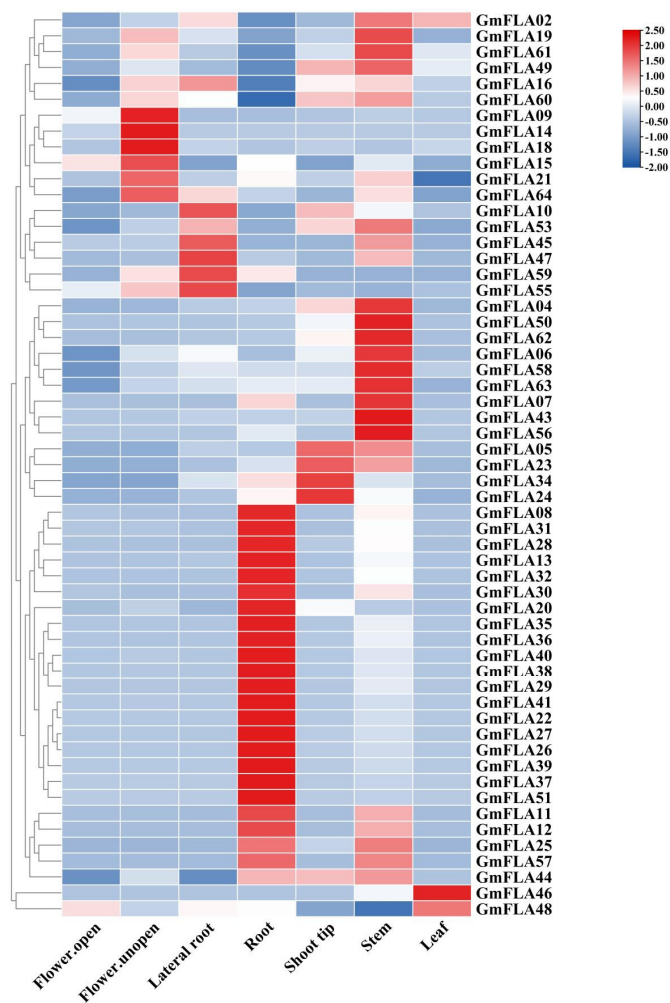


Figure 6: Tissue-specific expression patterns of *GmFLA* genes across seven soybean tissues (root standard, lateral root, stem, leaf, shoot tip, open flower, unopened flower). Expression values are presented as FPKM (fragments per kilobase of transcript per million mapped reads) derived from public RNA-seq data (Phytozome v13.1). Color scale from blue to red indicates low to high expression levels. Hierarchical clustering (Euclidean distance, complete linkage) groups genes with similar expression patterns. Genes with no detectable expression (FPKM = 0) are shown in white.

3.8 Gene Expression Pattern Analysis

Nine *GmFLA* genes exhibiting high expression levels specifically in stems were selected to investigate their expression patterns across six distinct soybean varieties (Fig. 7). The varieties JSL017, JSC041, and JSL079 possess thick stems (average diameter > 11 mm), whereas JSC028, JSL094, and JSL108 exhibit slender stems (average stem diameter < 6 mm). Expression levels of these genes varied significantly among varieties. For instance, *GmFLA47* showed the highest expression in JSL017 but the lowest in JSL108 (Fig. 7A); conversely, *GmFLA62* exhibited peak expression in JSL094 and the lowest level in JSL017 (Fig. 7E). In contrast, the expression levels of *GmFLA58*, *GmFLA61*, and *GmFLA19* did not differ significantly between the thick-stemmed and slender-stemmed groups (Fig. 7B–D). Notably, *GmFLA62*, *GmFLA43*, *GmFLA50*, *GmFLA07*, and *GmFLA56* displayed consistently lower expression in the thick-stemmed varieties compared with the slender-stemmed ones with particularly pronounced reductions observed for *GmFLA07* and *GmFLA56* (Fig. 7E–I). Overall, *GmFLA07*, *GmFLA56*, and *GmFLA62* showed significant lower expression between in the thick-stemmed group compared to the slender-stemmed group (Fig. 7E,H,I). No phenotypic or environmental association is inferred from this small-scale expression survey.

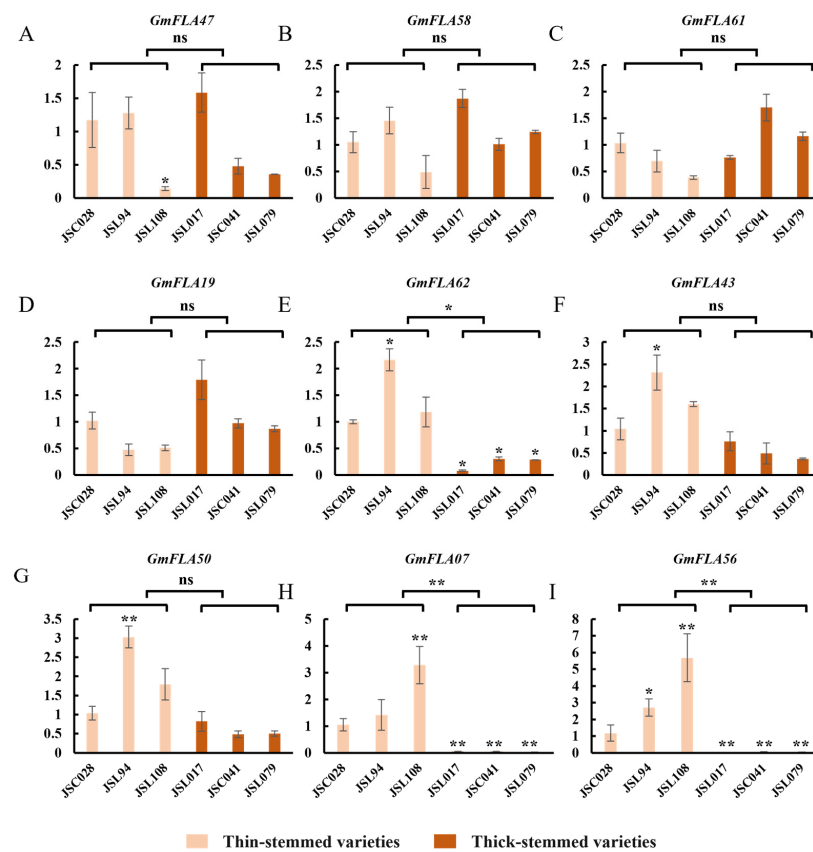


Figure 7: Expression levels of nine stem-preferential *GmFLA* genes in six soybean varieties differing in stem thickness. Varieties are classified as thick-stemmed (JSL017, JSC041, JSL079) and thin-stemmed (JSC028, JSL094, JSL108). Relative mRNA levels were determined by qPCR and are presented as mean + SD of three biological replicates, each with three technical replicates ($n = 3$ independent experiments). The genes presented in the panels are: A (*GmFLA47*), B (*GmFLA58*), C (*GmFLA61*), D (*GmFLA19*), E (*GmFLA62*), F (*GmFLA43*), G (*GmFLA50*), H (*GmFLA07*), and I (*GmFLA56*). Asterisks indicate significant differences compared with the thin-stemmed variety JSC028 or between thin and thick-stemmed groups, with ns ($p > 0.05$), * ($p < 0.05$), and ** ($p < 0.01$), respectively.

4 Discussion

FLA is one class of arabinogalactan proteins that have significant impact on plant growth and development, especially on secondary cell wall biosynthesis and stress responses [24,25]. The FLA gene family has been identified in some plant species; however, information is still lacking in legumes, especially soybean. This study reports the identification 64 FLA genes in soybean and presents a detailed analysis encompassing their domain architecture, phylogenetic tree, gene structure, conserved motifs, and expression patterns.

We grouped the *GmFLA* into different groups according to the evolutionary tree and the domain characters. Most of the members of group A have a single fasciclin domain with a C-terminal GPI anchor. These structural features are similar to those of orthologous genes in other plant species, where they are involved in various development processes [11,12]. The presence of the GPI anchor targets these proteins to the outer surface of the plasma membrane [19,36], positioning group A *GmFLA* proteins at the cell surface. This subcellular location indicates that they may function as receptors or adhesion molecules, contributing to cell wall stabilization and facilitating cell-to-cell communicate. Such localization is critical for protein function, enabling interactions with other cell surface proteins and impacting signaling pathways that regulate important physiological processes. Understanding these dynamics provides insight into the evolution of plant proteins and their developmental roles [27]. GPI-anchored proteins localized to the cell wall commonly function as surface receptors or cell adhesive molecules [4]. They play a vital role in regulating the cell wall and the signaling pathways which accompany it.

Previous studies have reported that orthologous of group A members, such as *AtFLA11* and *AtFLA12* in *Arabidopsis*, predominantly function in regulating secondary wall formation and mediating responses to mechanical forces, thereby influencing cell wall strength, rigidity and maintenance [28]. Notably, we identified exceptions within Group A characterized by two fasciclin domains and a GPI anchor, suggesting specialized functional roles distinct from the canonical single-domain members. In contrast, group B exhibits considerable structural and domain diversity, including a consistent absence of GPI anchors, which indicates functional divergence. For instance, *GmFLA59* contains one fasciclin domain and is predicted to be localized to the nucleus. Given that fasciclin domains are known to mediate cell adhesion and intercellular communication in other organisms [38], we hypothesize that *GmFLA59* may play similar roles in soybean [39]. The lack of GPI anchors in Group B suggests that these proteins may function intracellularly or be released without membrane anchoring, potentially engaging in nuclear signaling pathways rather than direct modification of the cell wall.

Group C, the smallest group, is characterized by a conserved architecture: all members possess two fasciclin domains, a GPI anchor, and two or more AGP glycomodule regions. This consistent domain structure suggests these proteins fulfill highly specific biological roles. The presence of two fasciclin domains is known to facilitate cell-cell adhesion and cell-extracellular matrix interactions [39]. The conserved two-domain structure in Group C *GmFLA* proteins indicates a specialized role in cell adhesion, potentially in the formation of the secondary cell wall in stems. *GmFLA55* represents a notable exception within this group, possessing only a single fasciclin domain and lacking a GPI anchor. This distinct structure implies a divergent function, potentially involving extracellular activity, modification of the extracellular matrix, or facilitation of intercellular communication.

In the present study, a total of 55 duplication events were detected among the 64 *GmFLA* genes, including 50 segmental duplication events and 5 tandem duplication events, indicating that gene duplication contributed extensively to the expansion of the *FLA* gene family in soybean. Consistent with the evolutionary pattern of FLAs across the plant kingdom reported by He et al. [40], both segmental and tandem duplications

served as important driving forces for the amplification of *GmFLA* genes. Notably, segmental duplication was overwhelmingly predominant compared with tandem duplication in soybean, which is highly consistent with the findings in maize [41], in which segmental duplication was the major contributor to *FLA* family expansion while tandem duplication played only a minor role. Collectively, these observations suggest that segmental duplication may represent a conserved and dominant mechanism underlying the expansion of the *FLA* gene family in plants, whereas tandem duplication acts as a complementary force in a species-specific manner. All the K_a/K_s ratios of the paired duplicated genes were less than 1, suggesting strong purifying selection (Table S5). This finding has significant evolutionary implications: it suggests that *GmFLA* genes have preserved essential functions over millions of years of diversification, and that even minor changes in these proteins could be deleterious to plant fitness. The absence of positive selection ($K_a/K_s > 1$) suggests that any neofunctionalization that may have occurred likely involved changes in gene expression patterns rather than modifications to protein-coding sequences [41].

In this study, a total of 20 conserved motifs were identified in the 64 *GmFLA* proteins. The conserved arrangement of these motifs provides important clues for understanding functional divergence and evolutionary relationships within the *FLA* family. The distinct motif compositions observed among different phylogenetic subgroups strongly suggest subclass-specific functional specialization. Notably, motif 2 was uniquely present in Group A, which is consistent with findings in maize that subgroup-specific motifs are closely associated with stress responses and developmental regulation [41]. As most Group A *GmFLA* genes are putatively involved in secondary cell wall biosynthesis, motif 2 may contribute critically to cell wall stability and mechanical strength. Groups A and B shared motifs 14 and 5, implying partially conserved functions between these two subgroups. This pattern resembles that reported in maize, in which shared motifs across phylogenetic clades correspond to conserved biological functions despite sequence divergence [40]. Group C exhibited a distinct motif signature characterized by motifs 4 and 13, and nearly all members contained two fasciclin domains and a GPI anchor. Previous studies indicated that FLAs with two fasciclin domains participate in pollen tube growth and stress-related cell growth control [40,42]. Accordingly, the unique motif architecture of Group C *GmFLAs* implies specialized functions, likely in secondary cell wall assembly during stem development.

FLA proteins typically contain [A/S/T]-Pro-rich glycomodules known as AGP regions, which are generally defined by the consensus motif [A/S/T]-P-X(0–10)-[A/S/T]-P or at least two to four consecutive [A/S/T]-P repeats, outside the FAS domain and N/C-terminal signal peptides [11,21]. By definition, FLAs are distinguished by the simultaneous presence of a fasciclin (FAS) domain and AGP-like regions. However, in the present study, *GmFLA01* was identified as an *FLA* despite lacking a typical AGP region. This is not unprecedented, as previous studies have reported FAS domain-containing proteins that lack AGP regions but are still classified as FLAs based on overall sequence homology and phylogenetic placement within the *FLA* clade [21]. Consistently, our phylogenetic tree showed that *GmFLA01* clustered firmly within Group A together with other *GmFLAs* harboring AGP regions (Fig. 1), supporting its classification as a bona fide *FLA*. Future investigation into proline O-glycosylation status in *GmFLA01* would help clarify its functional properties.

In addition, several *GmFLAs* exhibited incomplete conserved domains. For instance, *GmFLA54* lacked the conserved H2 domain, a pattern also reported in FLAs from jute and *Nicotiana* [42,43]. Moreover, five *GmFLAs* lacked the [Y/F]H motif, which in rice has been implicated in carbohydrate binding and cell wall interactions [38]. The absence of these key motifs in specific *GmFLAs* implies potential functional divergence, modified molecular mechanisms, or distinct interacting partners during cell wall-associated processes.

Promoter analysis revealed that *GmFLA* genes are enriched in hormone-, light- and stress-responsive cis-elements, especially those related to MeJA and ABA signaling, indicating that these genes may integrate developmental and stress responses to coordinate cell wall remodeling with environmental adaptation. Such regulatory patterns are consistent with the conserved roles of FLAs in *Populus*, maize and *Nicotiana*, in which ABA and MeJA responsiveness mediates stress-induced cell wall modification [25,41,43]. This regulatory link between *GmFLA* genes, cell wall remodeling, and stress signaling aligns with recent findings on soybean defense against charcoal rot (*Macrophomina phaseolina*), a major soil-borne pathogen targeting cell wall integrity. Khalili et al. noted that β -glucosidase is key for *M. phaseolina* management by regulating cell wall hydrolysis and defense activation [44]. As core cell wall glycoproteins involved in structural stability and stress signaling, certain *GmFLAs* (e.g., stem/root-expressed members) may interact with β -glucosidase pathways or enhance *Macrophomina* resistance by reinforcing cell wall barriers and modulating early defense.

Tissue-specific expression profiling further supported functional diversification within the *GmFLA* family. Approximately 37.5% of *GmFLAs* were preferentially expressed in roots, consistent with the known functions of *AtFLA1* in *Arabidopsis* root tip development and stem cell maintenance [18], implying conserved roles in soybean root architecture and nutrient acquisition [15,17]. Six *GmFLAs* were highly expressed in unopened flowers, suggesting potential functions in pollen development, ovule formation or pollination, processes closely linked to yield determination [38,45]. One-third of *GmFLAs* showed stem-enriched expression, a pattern widely conserved across plants including *Arabidopsis*, eucalyptus and poplar, in which FLAs determine secondary cell wall properties and stem mechanical strength [24,25,27,30]. These stem-preferential *GmFLAs* thus represent strong candidates for improving lodging resistance in soybean. By contrast, only two *GmFLAs* were leaf-specific, potentially contributing to vascular development and leaf architecture, similar to the roles of *OsFLA02/09* in regulating rice leaf angle [20].

The critical roles of *FLAs* in cell wall formation and plant development—and their tight spatiotemporal regulation during specific developmental stages and in response to environmental signals—have been well established [6,9,10]. In this study, we observed substantial variation in the expression levels of different *GmFLA* genes across soybean varieties, suggesting the existence of additional, genotype-specific regulatory mechanisms within this family. Indicatively, expression of *GmFLA47* was more in JSL094 but less in JSL108 whereas *GmFLA56* and *GmFLA07* exhibited the converse. Experimentally no upstream regulators of *FLA* genes have been validated as yet. Determining possible regulatory factors e.g., by co-expression network analysis give an important basis to further characterize the functions of this gene family [43,46]. Moreover, the expression levels of *GmFLA47*, *GmFLA58*, *GmFLA61*, and *GmFLA19* did not differ significantly between thick- and thin-stemmed plants, suggesting they do not directly determine stem thickness. Instead, they may contribute to other processes such as maintaining cell wall integrity, cell–cell adhesion, or responses to environmental stimuli. Due to the limited sample size (only involving six varieties), no conclusive association between gene expression and stem thickness can be drawn; for such research to be conducted, verification is needed in a larger and controlled sample group. Understanding their precise functions in future will deepen our understanding of plants stress adaptation and offer potential targets for improving soybean productivity [47,48].

The findings presented in this study are primarily descriptive and correlational. While expression analyses provide insights into the putative functions of *GmFLA* proteins, definitive functional roles cannot be made without the experimental validation [21,29,49]. The association analysis between *GmFLA* expression and stem thickness was carried out using six varieties; an extended germplasm panel would strengthen

the reported correlations [2,3]. Additionally, the RNA-seq expression data were taken through publicly available datasets and should be validated experimentally.

Future studies should employ genetic approaches to test the hypothesized roles of *GmFLA* genes, moving beyond characterization. Suggested methods include: (1) CRISPR-Cas9-mediated mutagenesis of candidate genes (e.g., *GmFLA07* or *GmFLA56*) to assess their effects on stem thickness and mechanical strength; (2) overexpression experiments in soybean or heterologous systems to evaluate gain-of-function phenotypes; and (3) complementation assays in *Arabidopsis* mutants [27,28,30]. These experimental approaches are essential to establish whether the observed correlative relationships reflect direct functional contributions to stem development.

5 Conclusions

The study gives the initial systemic genomic, evolutionary, and expression-based characterization of the FLA gene family in soybean. It determines 64 *GmFLA* genes, measures their expansion via segmental duplication under purifying selection, characterizes their tissue-specific expression, and shows correlational evidence linking five genes (including *GmFLA07* and *GmFLA56*) to stem thickness, which is directly related to lodging resistance. This study extends previous FLA investigation in other crop species by establishing, for the first time, an association between FLA gene expression and soybean stem thickness, showing its novelty and agronomic relevance. Although causal functions are yet to be experimentally validated, the candidate genes and quantitative expression controls developed here offer a resource base to be used in the future functional studies that seek to understand and possibly enhance stem thickness in soybean and other legume crops. This study is limited in various ways. The results presented here are mainly descriptive and correlational. There were no functional experiments (e.g., CRISPR-Cas9 mutagenesis, overexpression or complementation assays). Thus, we found some *GmFLA* genes to be differentially expressed in the thick and thin-stem soybean varieties, these results do not imply that any *GmFLA* gene play a causal role in influencing stem development, stem thickness, or lodging resistance. Further research using reverse genetics methods is needed to determine the performance of these genes in stem growth. Hence, the present research provides a foundation for integrating genome-wide identification gene family analysis with agronomic attributes improvement in soybean plants.

Acknowledgement: Not applicable.

Funding Statement: The authors received no specific funding for this study.

Author Contributions: Conceptualization, Mazin Ahmed Abdelraouf and Li Song; Methodology, Hind Abdelmonim Elsanosi; Software, Hind Abdelmonim Elsanosi; Validation, Mazin Ahmed Abdelraouf, Xiaoqi He, Tiantian Zhu, Jinghui Shi, and Ullah Habib; Formal Analysis, Mazin Ahmed Abdelraouf; Investigation, Xiaoqi He, Tiantian Zhu, Jinghui Shi, and Ullah Habib; Data Curation, Mazin Ahmed Abdelraouf; Writing—Original Draft Preparation, Mazin Ahmed Abdelraouf and Li Song; Writing—Review and Editing, Li Song; Visualization, Mazin Ahmed Abdelraouf; Supervision, Li Song; Project Administration and Funding Acquisition, Li Song All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: All data generated or analyzed during this study are included in this published article and its Supplementary Materials.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Supplementary Materials: The supplementary material is available online at <https://www.techscience.com/doi/10.32604/phyton.2026.079749/s1>. Table S1. Gene-specific primers used for quantitative real-time PCR (qRT-PCR) analysis of selected *GmFLA* genes. Table S2. Domain architecture, GPI anchor status, and AGP region counts of 64 *GmFLA* proteins. Note: “+” indicates presence of GPI anchor signal; “-” indicates absence. “≥2” indicates two or more AGP glycomodule regions. Table S3. Characteristics of 64 *GmFLA* proteins, including gene identifiers, protein properties, subcellular localization, domain architecture, GPI anchor status, and AGP region counts. Table S4. The events of *GmFLA* duplication. Table S5. Non-synonymous (Ka) and synonymous (Ks) substitution rates for segmentally duplicated *GmFLA* gene pairs. Ka/Ks ratios < 1 indicate purifying selection; ratios > 1 indicate positive selection. Table S6. Cis-regulatory elements identified in the promoter regions (1500 bp upstream of ATG) of *GmFLA* genes. Elements were identified using PlantCARE. Coordinates (Start, End) refer to positions relative to the translation start site (ATG = position 0). Table S7. Expression profiles of 64 *GmFLA* genes across seven soybean tissues. FPKM (fragments per kilobase of transcript per million mapped reads) values were obtained from RNA-seq data. Tissues: Flower.open (open flower), Flower.unopen (unopened flower), Lateral root, Root, Shoot tip, Stem, Leaf.

Abbreviations

FLA	protein fasciclin-like arabinogalactan proteins
AGPs	Arabinogalactan proteins
GRPs	Glycine-rich proteins
HRGPs	Hydroxyproline-rich glycoproteins
pI	Isoelectric point
MW	Molecular weight
GRAVY	Grand average of hydropathicity
MeJA	Methyl Jasmonate
aa	amino acids
GPI	glycosyl phosphatidylinositol
MEME	Multiple EM for Motif Elicitation
FPKM	fragments per kilobase of transcript per million mapped reads
Ka	non-synonymous substitution rate
Ks	synonymous substitution rate
WGD	whole-genome duplication
NCBI	National Center for Biotechnology Information
BIC	Bayesian Information Criterion
NJ	neighbor-joining
bp	base pair
SignalP	signal peptide prediction tool

References

- Pagano MC, Miransari M. Abiotic and biotic stresses in soybean production. Amsterdam, The Netherlands: Elsevier; 2016. p. 1–26. [[CrossRef](#)].
- Zhao W, Zeng D, Zhao C, Han D, Li S, Wen M, et al. Identification of QTLs and key genes enhancing lodging resistance in soybean through chemical and physical trait analysis. *Plants*. 2024;13(24):3470. [[CrossRef](#)].
- Xu Z, Zhang L, Kong K, Kong J, Ji R, Liu Y, et al. Creeping Stem 1 regulates directional auxin transport for lodging resistance in soybean. *Plant Biotechnol J*. 2025;23(2):377–94. [[CrossRef](#)].
- Wang C, Huang S, Zhang X, Shan F, Fan J, Lyu X, et al. Red and blue light differentially regulate carbon metabolism and stem elongation in soybean cultivars with contrasting lodging resistance under shade: A multi-omics perspective. *J Photochem Photobiol B Biol*. 2025;271:113241. [[CrossRef](#)].
- Jiang Y, Balasubramanian B, Park S, Anand A, Meyyazhagan A, Pappusamy M, et al. Green nanoparticles in agriculture: Enhancing crop growth and stress tolerance. *Plant Stress*. 2025;18:101017. [[CrossRef](#)].
- Ellis C, Karafyllidis I, Wasternack C, Turner JG. The *Arabidopsis mutant cev1* links cell wall signaling to jasmonate and ethylene responses. *Plant Cell*. 2002;14(7):1557–66. [[CrossRef](#)].

7. Xie D, Ma L, Šamaj J, Xu C. Immunohistochemical analysis of cell wall hydroxyproline-rich glycoproteins in the roots of resistant and susceptible wax gourd cultivars in response to *Fusarium oxysporum* f. sp. *benincasae* infection and fusaric acid treatment. *Plant Cell Rep.* 2011;30(8):1555–69. [[CrossRef](#)].
8. Niu Y, Hu B, Li X, Chen H, Takáč T, Šamaj J, et al. Comparative digital gene expression analysis of tissue-cultured plantlets of highly resistant and susceptible banana cultivars in response to *Fusarium oxysporum*. *Int J Mol Sci.* 2018;19(2):350. [[CrossRef](#)].
9. Mashiguchi K, Urakami E, Hasegawa M, Sanmiya K, Matsumoto I, Yamaguchi I, et al. Defense-related signaling by interaction of Arabinogalactan proteins and β -glucosyl yariv reagent inhibits gibberellin signaling in barley aleurone cells. *Plant Cell Physiol.* 2008;49(2):178–90. [[CrossRef](#)].
10. Van Holle S, Van Damme EJM. Signaling through plant lectins: Modulation of plant immunity and beyond. *Biochem Soc Trans.* 2018;46(2):217–33. [[CrossRef](#)].
11. Showalter AM, Keppler B, Lichtenberg J, Gu D, Welch LR. A bioinformatics approach to the identification, classification, and analysis of hydroxyproline-rich glycoproteins. *Plant Physiol.* 2010;153(2):485–513. [[CrossRef](#)].
12. Showalter AM. Arabinogalactan-proteins: Structure, expression and function. *Cell Mol Life Sci CMLS.* 2001;58(10):1399–417. [[CrossRef](#)].
13. Ellis M, Egelund J, Schultz CJ, Bacic A. Arabinogalactan-proteins: Key regulators at the cell surface? *Plant Physiol.* 2010;153(2):403–19. [[CrossRef](#)].
14. Ma Y, Yan C, Li H, Wu W, Liu Y, Wang Y, et al. Bioinformatics prediction and evolution analysis of Arabinogalactan proteins in the plant kingdom. *Front Plant Sci.* 2017;8:66. [[CrossRef](#)].
15. Leszczuk A, Kalaitzis P, Kulik J, Zdunek A. Review: Structure and modifications of Arabinogalactan proteins (AGPs). *BMC Plant Biol.* 2023;23(1):45. [[CrossRef](#)].
16. Zhang Q, Gong M, Xu X, Li H, Deng W. Roles of auxin in the growth, development, and stress tolerance of horticultural plants. *Cells.* 2022;11(17):2761. [[CrossRef](#)].
17. Johnson KL, Jones BJ, Bacic A, Schultz CJ. The fasciclin-like Arabinogalactan proteins of *Arabidopsis*. A multigene family of putative cell adhesion molecules. *Plant Physiol.* 2003;133(4):1911–25. [[CrossRef](#)].
18. Johnson KL, Kibble NAJ, Bacic A, Schultz CJ. A fasciclin-like Arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, *fla1*, shows defects in shoot regeneration. *PLoS One.* 2011;6(9):e25154. [[CrossRef](#)].
19. Eisenhaber B, Wildpaner M, Schultz CJ, Borner GH, Dupree P, Eisenhaber F. Glycosylphosphatidylinositol lipid anchoring of plant proteins. sensitive prediction from sequence- and genome-wide studies for *Arabidopsis* and rice. *Plant Physiol.* 2003;133(4):1691–701. [[CrossRef](#)].
20. Liang R, You L, Dong F, Zhao X, Zhao J. Identification of hydroxyproline-containing proteins and hydroxylation of proline residues in rice. *Front Plant Sci.* 2020;11:1207. [[CrossRef](#)].
21. Huang GQ, Xu WL, Gong SY, Li B, Wang XL, Xu D, et al. Characterization of 19 novel cotton *FLA* genes and their expression profiling in fiber development and in response to phytohormones and salt stress. *Physiol Plant.* 2008;134(2):348–59. [[CrossRef](#)].
22. Li J, Wu X. Genome-wide identification, classification and expression analysis of genes encoding putative fasciclin-like Arabinogalactan proteins in Chinese cabbage (*Brassica rapa* L.). *Mol Biol Rep.* 2012;39(12):10541–55. [[CrossRef](#)].
23. MacMillan CP, Taylor L, Bi Y, Southerton SG, Evans R, Spokevicius A. The fasciclin-like Arabinogalactan protein family of *Eucalyptus grandis* contains members that impact wood biology and biomechanics. *New Phytol.* 2015;206(4):1314–27. [[CrossRef](#)].
24. MacMillan CP, Mansfield SD, Stachurski ZH, Evans R, Southerton SG. Fasciclin-like Arabinogalactan proteins: Specialization for stem biomechanics and cell wall architecture in *Arabidopsis* and *Eucalyptus*. *Plant J.* 2010;62(4):689–703. [[CrossRef](#)].
25. Zang L, Zheng T, Chu Y, Ding C, Zhang W, Huang Q, et al. Genome-wide analysis of the fasciclin-like Arabinogalactan protein gene family reveals differential expression patterns, localization, and salt stress response in *Populus*. *Front Plant Sci.* 2015;6:1140. [[CrossRef](#)].
26. Guerriero G, Mangeot-Peter L, Legay S, Behr M, Lutts S, Siddiqui KS, et al. Identification of fasciclin-like Arabinogalactan proteins in textile hemp (*Cannabis sativa* L.): *In silico* analyses and gene expression patterns in different tissues. *BMC Genom.* 2017;18(1):741. [[CrossRef](#)].

27. Shi H, Kim Y, Guo Y, Stevenson B, Zhu JK. The Arabidopsis *SOS5* Locus encodes a putative cell surface adhesion protein and is required for normal cell expansion. *Plant Cell*. 2003;15(1):19–32. [[CrossRef](#)].
28. Ma Y, MacMillan CP, de Vries L, Mansfield SD, Hao P, Ratcliffe J, et al. FLA11 and FLA12 glycoproteins fine-tune stem secondary wall properties in response to mechanical stresses. *New Phytol*. 2022;233(4):1750–67. [[CrossRef](#)].
29. Liu E, MacMillan CP, Shafee T, Ma Y, Ratcliffe J, van de Meene A, et al. Fasciclin-like Arabinogalactan-protein 16 (FLA16) is required for stem development in *Arabidopsis*. *Front Plant Sci*. 2020;11:615392. [[CrossRef](#)].
30. Wang H, Jin Y, Wang C, Li B, Jiang C, Sun Z, et al. Fasciclin-like arabinogalactan proteins, PtFLAs, play important roles in GA-mediated tension wood formation in *Populus*. *Sci Rep*. 2017;7(1):6182. [[CrossRef](#)].
31. Tamura K, Stecher G, Kumar S. MEGA11: Molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7. [[CrossRef](#)].
32. Zhou T, Xu K, Zhao F, Liu W, Li L, Hua Z, et al. Itol.toolkit accelerates working with iTOL (Interactive Tree of Life) by an automated generation of annotation files. *Bioinformatics*. 2023;39(6):btad339. [[CrossRef](#)].
33. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2—A multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91. [[CrossRef](#)].
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. [[CrossRef](#)].
35. Bailey TL, Elkan C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology; 1994 Aug 14–17; Menlo Park, CA, USA. p. 28–36.
36. Borner GH, Antrobus R, Hirst J, Bhumbra GS, Kozik P, Jackson LP, et al. Multivariate proteomic profiling identifies novel accessory proteins of coated vesicles. *J Cell Biol*. 2012;197(1):141–60. [[CrossRef](#)].
37. Zhu T, He X, Shi J, Zhang W, Chen H, Song L. Identification and expression analysis of soybean GmROP gene family. *J Yangzhou Univ*. 2025; 46(3):1671–4652. (In Chinese).
38. Pereira AM, Pereira LG, Coimbra S. Arabinogalactan proteins: Rising attention from plant biologists. *Plant Reprod*. 2015;28(1):1–15. [[CrossRef](#)].
39. Huber O, Sumper M. Algal-CAMs: Isoforms of a cell adhesion molecule in embryos of the *Alga volvox* with homology to *Drosophila* fasciclin I. *EMBO J*. 1994;13(18):4212–22. [[CrossRef](#)].
40. He J, Zhao H, Cheng Z, Ke Y, Liu J, Ma H. Evolution analysis of the fasciclin-like Arabinogalactan proteins in plants shows variable fasciclin-AGP domain constitutions. *Int J Mol Sci*. 2019;20(8):1945. [[CrossRef](#)].
41. Cao Y, Chen X, Wei B, Zeng T, Wang H, Zhu B, et al. Genome-wide identification and analysis of the FLA gene family in maize (*Zea mays* L.) and its expression in response to abiotic stresses. *BMC Plant Biol*. 2025;25(1):982. [[CrossRef](#)].
42. Hossain MS, Ahmed B, Ullah MW, Aktar N, Haque MS, Islam MS. Genome-wide identification of fasciclin-like Arabinogalactan proteins in jute and their expression pattern during fiber formation. *Mol Biol Rep*. 2020;47(10):7815–29. [[CrossRef](#)].
43. Wu X, Lai Y, Lv L, Ji M, Han K, Yan D, et al. Fasciclin-like Arabinogalactan gene family in *Nicotiana benthamiana*: Genome-wide identification, classification and expression in response to pathogens. *BMC Plant Biol*. 2020;20(1):305. [[CrossRef](#)].
44. Khalili E, Kamyab H, Balasubramanian B, Chelliapan S. Integrated management of charcoal rot (*Macrophomina phaseolina*) in soybean: Current strategies and the emerging role of β -glucosidase. *Appl Soil Ecol*. 2026;217:106549. [[CrossRef](#)].
45. Costa M, Pereira AM, Pinto SC, Silva J, Pereira LG, Coimbra S. *In silico* and expression analyses of fasciclin-like Arabinogalactan proteins reveal functional conservation during embryo and seed development. *Plant Reprod*. 2019;32(4):353–70. [[CrossRef](#)].
46. Yao K, Yao Y, Ding Z, Pan X, Zheng Y, Huang Y, et al. Characterization of the FLA gene family in tomato (*Solanum lycopersicum* L.) and the expression analysis of SFLAs in response to hormone and Abiotic Stresses. *Int J Mol Sci*. 2023; 24(22):16063. [[CrossRef](#)].
47. Bygdell J, Srivastava V, Obudulu O, Srivastava MK, Nilsson R, Sundberg B, et al. Protein expression in tension wood formation monitored at high tissue resolution in *Populus*. *J Exp Bot*. 2017;68(13):3405–17. [[CrossRef](#)].

48. Dahiya P, Findlay K, Roberts K, McCann MC. A fasciclin-domain containing gene, *ZeFLA11*, is expressed exclusively in xylem elements that have reticulate wall thickenings in the stem vascular system of *Zinnia elegans* cv Envy. *Planta*. 2006;223(6):1281–91. [[CrossRef](#)].
49. Seifert GJ, Roberts K. The biology of Arabinogalactan proteins. *Annu Rev Plant Biol*. 2007;58:137–61. [[CrossRef](#)].