Plant Tissue Cult. & Biotech. **34**(2): 217-235, 2024 (December) **ISSN 1817-3721, E-ISSN 1818-8745**
DOI: https://doi.org/10.3329/ptcb.v34i2.78832

PTC&B

Genome-Wide Identification and *In silico* Expression Analysis of Serine Carboxypeptidase-Like (SCPL) Proteins in Soybean (*Glycine max* L.)

Faisal Ahmed, Mst. Tamima Sultana[1], Niloy Chowdhury, Bivas Kumar Sarkar, G. M. Al Amin, Tahmina Islam[1] and Ripa Akter Sharmin*

Department of Botany, Jagannath University, Dhaka-1100, Bangladesh

Key words: Serine carboxypeptidases-like protein, Genome-wide analysis, Soybean, Gene expression

Abstract

The serine carboxypeptidase-like (SCPL) protein family is crucial in plant development and stress response regulation. However, information on SCPL proteins in soybean remains limited. Here, we identified 84 *GmSCPL* genes distributed across 19 chromosomes. Phylogenetic analysis grouped *GmSCPLs* into three clusters, closely related to *SCPLs* in *Arabidopsis thaliana* and *Oryza sativa*. Duplication analysis revealed 31 duplicated gene pairs, with 27 linked to whole-genome or segmental duplication and 4 to tandem duplication. Three hub genes, *GmSCPL82*, *GmSCPL46*, and *GmSCPL16* were identified through protein-protein interaction analysis. Promoter regions predominantly contain *cis*-regulatory elements for light, hormone, environmental, and growth responses. RNA-seq data highlighted *GmSCPL20*, *GmSCPL33*, and *GmSCPL76* as highly up-regulated under heat stress; *GmSCPL40*, *GmSCPL64*, and *GmSCPL69* under drought; and *GmSCPL2*, *GmSCPL27*, and *GmSCPL37* under salt stress. This study conducted a comprehensive analysis of the SCPL protein family in soybeans, offering a valuable foundation for future research on the role of *SCPL* genes in abiotic stress tolerance.

Introduction

Serine carboxypeptidase-like (SCPL) proteins constitute a vital family of enzymes that catalyze the hydrolysis of C-terminal peptide bonds in proteins and peptides (Jiang et al. 2020). These proteins are defined by a tertiary α/β hydrolase structure, featuring a unique catalytic center and a conserved topology (Xu et al. 2021). Structurally, SCPL proteins include a signal peptide for cellular signaling and transport, four substrate-binding sites,

*Author for correspondence: <ripasharmin@bot.jnu.ac.bd>. [1]Plant Breeding and Biotechnology Laboratory, Department of Botany, University of Dhaka, Dhaka-1000, Bangladesh.

several N-glycosylation sites, and evolutionarily conserved regions critical to their catalytic role (Milkowski and Strack 2004). Although SCPL proteins primarily act as peptidases, certain members of this family also display acyltransferase activity, utilizing the O-glucose acyl group (1-O-β-glucose ester) as an acylation substrate (Agarwal et al. 2012, Ahmad et al. 2020).

In recent years, substantial efforts have focused on uncovering the roles of *SCPL* genes in various plant species. The expression of *OsBISCPL1* in rice is induced by jasmonic acid and salicylic acid and is up-regulated in response to blast fungus infection. Transgenic plants expressing *OsBISCPL1* show reduced sensitivity to abscisic acid and partial resistance to oxidative stress (Liu et al. 2008). Additionally, the protein brassinosteroid insensitive 1 is linked to sinapate ester biosynthesis, suggesting that SCPL proteins may participate in essential biochemical pathways supporting plant growth and resilience under environmental stress (Milkowski and Strack 2004). In *Arabidopsis*, plants overexpressing the extra carpels and seeds gene displayed sensitivity to brassinolide (Wen et al. 2012), while *sad7* gene expression was associated with enhanced resistance to soil-borne pathogens. SCPL proteins are further implicated in responses to various abiotic stresses, including drought, salinity, light, and nutrient deficiencies (Ciarkowska et al. 2016, Wang et al. 2021). Consequently, exploring stress-related genes and understanding their functions holds great importance for developing stress-tolerant soybean varieties.

Soybean is a major source of vegetable protein and contains valuable human health-promoting phytochemicals, such as isoflavones and phenolic compounds. As a result, it ranks as the world's second-largest source of vegetable oil (Miyake et al. 2018, Qin et al. 2022). However, soybean production faces substantial challenges from both biotic and abiotic stressors, which negatively impact plant yield, and quality (Cohen et al. 2021, Hartman et al. 2011). Genome-wide analysis of the SCPL protein family has been carried out in various crops such as 54 in *Arabidopsis* (Fraser et al. 2005), 71 putative *SCPL* genes identified in rice (Feng and Xue 2006), 57 in poplar (Zhu et al. 2018), 47 in tea plant (Ahmad et al. 2020), 59 in grape (Wang et al. 2021), 210 in wheat (Xu et al. 2021), 117 in brassica (Liu et al. 2022), 96 in cotton (Wang et al. 2022), and 73 in soybean (He et al. 2024).

The availability of soybean genomic data has greatly facilitated the identification and functional characterization of SCPL proteins within a very short period. In this study, we conducted a comprehensive genome-wide analysis of the *SCPL* gene family in soybean, identifying a total of 84 *SCPL* genes. To gain insights into the evolution and function of these genes, we performed phylogenetic analysis, mapped their physical locations across chromosomes, analyzed gene structure, and assessed *in silico* gene expression patterns using RNA-seq data. The findings from this study enhance our understanding of *SCPL* gene evolution and their roles in regulating growth, development, and responses to abiotic stress in soybean.

Materials and Methods

To identify *SCPL* genes within the soybean genome, 51 SCPL protein sequences from *Arabidopsis thaliana* (Fraser et al. 2005) were first downloaded from the TAIR database (http://www.arabidopsis.org/). These *A. thaliana* SCPL protein sequences were used as queries in a BLASTP search against the Ensemble Plants database (https://plants.ensembl.org/index.html), with an E-value threshold of $\leq 1 \times 10^{-5}$. The resulting candidate SCPL proteins from soybean were further validated using HMMscan (https://www.ebi.ac.uk/Tools/hmmer/) to confirm the presence of the PF00450.25 domain, which corresponds to the peptidase_S10 or peptidase_S10 superfamily domain (CL0028) (Potter et al. 2018). Proteins containing this domain were identified as GmSCPL and were named sequentially based on their chromosomal location.

The ExPASy website was (https://www.expasy.org/) utilized to predict the physicochemical properties of SCPL proteins, including amino acid count, coding sequence (CDS) length, molecular weight (MW), and isoelectric point (pI) (Gasteiger et al. 2005). The grand averages of hydropathy (GRAVY) values were determined using the GRAVY CALCULATOR (http://www.gravy-calculator.de/). Subcellular localization predictions were conducted using the online tool CELLO (http://cello.life.nctu.edu.tw/) (Yu et al. 2006). To validate the predicted cellular locations obtained from CELLO, WoLF PSORT (https://wolfpsort.hgc.jp/) was also employed (Horton et al. 2007).

A Maximum-Likelihood (ML) tree was constructed to illustrate the phylogenetic relationships among the SCPL proteins (Tamura et al. 2011). The amino acid sequences of SCPL proteins from *Arabidopsis thaliana*, *Glycine max*, *Glycine soja*, and *Oryza sativa* were aligned using the MUSCLE algorithm in Molecular Evolutionary Genetics Analysis (MEGA) version X software (Kumar et al. 2018). Phylogenetic tree construction was performed in MEGA-X using the ML method with the Poisson substitution model, and the gaps/missing data were treated with the complete deletion option. The reliability of the tree was assessed using the bootstrap method with 1000 replicates (Felsenstein 1985). The XML file, Newick file of the evolutionary tree, and GFF file of the gene structure were processed and visualized using TBtools-II software. The final phylogenetic tree was edited and visualized through iTOL: Interactive Tree of Life (v.6) (https://itol.embl.de/) (Letunic and Bork 2021). Additionally, the MEME online tool (http://meme.nbcr.net/meme) was employed to identify conserved motifs within the GmSCPL proteins (Bailey et al. 2015). The exon/intron structure of the *SCPL* genes was analyzed using the Gene Structure Display Server (GSDS) (Hu et al. 2015). Domain analysis of the SCPL proteins was conducted with the NCBI Conserved Domain Search tool using default parameters. The domain organization was visualized using TBtools-II. (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The GTF file for the soybean genome was obtained from the Ensembl Plant database. Using the GTF file and gene IDs of the soybean SCPL family members, gene density and chromosomal maps were generated. TBtools-II software was employed to visualize the physical chromosome map and density profile, illustrating the positions of all *SCPL* genes (Chen et al. 2020).

Synonymous (Ks) and non-synonymous (Ka) substitution rates, along with Ka/Ks ratios, were calculated for two closely related *GmSCPL* genes using the Simple Ka/Ks Calculator in TBtools-II to assess gene duplication events and identify orthologous and paralogous gene pairs. Following the criteria of Wang et al. (2010), co-paralogs located on the same chromosome and within five or fewer genes in a 100 kb region were classified as tandem duplications, while those beyond this range were considered segmental duplications. The time of duplication events (Mya) was estimated using the formula T = Ks/(26.110^{-9})*10^{-6} million years ago (Mya) (Lynch and Conery 2000). A Ka/Ks ratio greater than 1 suggests positive selection, a ratio equal to 1 indicates neutral selection, and a ratio less than 1 implies negative or purifying selection (Hurst 2002). The collinear pairs of *Glycine max* were visualized using the Advanced Circos Option in TBtools-II.

Gene Ontology (GO) Enrichment Analysis was conducted to acquire GO annotations for soybean by inputting potential candidate protein or gene IDs into ShinyGO v0.80 (http://bioinformatics.sdstate.edu/go/). A false discovery rate (FDR) threshold of 0.01 was applied to determine the significance of GO enrichment (Mazhar et al. 2023). Search Tool for Retrieval of Interacting Genes/Proteins (STRING) online database was applied to construct the protein-protein interaction (PPI) network using *G. max* as background (Szklarczyk et al. 2019). Credible PPI interactions were further visualized with the network analyst 3.0 (Zhou et al. 2019).

The potential *cis*-acting regulatory elements (CAREs) were retrieved by examining the 2,000 bp upstream region of the start codon in the identified *GmSCPLs* using the Plant CARE database (http:// bioinformatics.psb.ugent.be/webtools/plantcare/html/) (Lescot et al. 2002). The visualization of these *cis*-regulatory elements was performed using TBtools-II software.

Publicly available RNA sequencing data for *G. max* under drought, heat stress (PRJNA644602), and salt stress (PRJNA432861) conditions were obtained from the NCBI Sequence Read Archive (SRA) (https://www.ncbi.nlm.nih.gov/sra/?term=). The datasets comprised three biological replicates for each treatment condition, focusing on leaf tissue samples. Paired-end reads were retrieved from the SRA database using the SRA Toolkit and converted to FASTQ format.

The resulting FASTQ files were split into individual read pairs to facilitate subsequent alignment using the SRA Toolkit (sratoolkit.3.1.1-win64). Read quality was assessed using FastQC, executed on the Galaxy web server (Afgan et al. 2018), to ensure data integrity before further processing. The HISAT2 tool, available on the same server, was employed to align the reads to the Glycine max reference genome. Following alignment, Cufflinks tools from the Galaxy server (Afgan et al. 2018), were used to calculate FPKM (fragments per kilobase of transcript per million mapped reads) values for each gene, providing a measure of gene expression levels across all three biological replicates for each condition. To visualize the expression patterns under various abiotic stresses, a heatmap was generated using TBtools-II (Chen et al. 2020). The mean

expression value was log2 transformed to create a heatmap that reflects different expression patterns resulting from drought, heat, and salt stress conditions.

Results and Discussion

After screening and validating the conserved domain, a total of 84 putative *GmSCPL* genes were identified. These genes were designated with unique identifiers, from *GmSCPL1* to *GmSCPL84,* based on their specific positions on the soybean chromosomes. Detailed information for each gene, including gene name, gene ID, chromosome number, gene location, strand (forward: + and reverse: -), protein size (aa), molecular weight (kDa), protein isoelectric point (pI) and predicted subcellular localization are identified (Table 1). The number of *SCPL* genes identified in soybean is like that reported in other species, though with some variation. For instance, the *Arabidopsis* genome contains 54 *SCPL* genes (Fraser et al. 2005), while the rice genome has 71 *SCPL* genes (Feng and Xue 2006). The higher number of *SCPL* genes in soybeans compared to *Arabidopsis* may be due to soybean's unique adaptation as a crop species and its more complex genome, which has undergone multiple rounds of polyploidization. In contrast, wheat exhibits a significantly greater number of *SCPL* genes, with 210 identified (Xu et al. 2021). Soybean SCPL proteins displayed significant variation in subcellular localization. Most of these proteins were predicted to be in various cellular organelles, including the extracellular, and lysosome (Table 1). Previous studies on SCPL proteins in *Arabidopsis* indicated their predominant localization in extracellular compartments (Fraser et al. 2005), while grape research revealed a primary localization in both extracellular and vacuolar compartments (Wang et al. 2021).

A comparative phylogenetic tree has been constructed to elucidate the evolutionary relationships of SCPL proteins in three species: *G. max* (84), *G. soja* (69), *A. thaliana* (51), and *O. sativa* (60), using all their SCPL members (Fig. 1). Phylogenetic analysis revealed that the 84 SCPL proteins in soybean clustered into three groups: Group I (violet), Group II (green), and Group III (orange), based on the evolutionary distance estimated. A larger number of proteins were distributed in Group I and Group II across the three species, with Group II forming the largest cluster containing 136 members, followed by Group I with 65 members, and Group III with 63 (Fig. 1). Previous studies have shown that SCPL members in rice and *Arabidopsis* similarly form three distinct groups, indicating evolutionary conservation across species (Feng and Xue 2006). In wheat, most SCPL proteins also fall into three categories, reflecting functional similarities between monocots and dicots (Xu et al. 2021). Notably, Group I in the soybean phylogeny, which included genes with potential acyltransferase activity, contained orthologs such as *AtSCPL8* and *AtSCPL10* from *Arabidopsis*, suggesting that *SCPL* genes in soybean may share functional similarities with their counterparts in other species. The conservation of acyltransferase activity within Group I across different species highlights the crucial role *SCPL* genes

Table 1. List of 84 putative *SCPL* genes identified in soybean with their molecular attributes.

Gene Name	Gene ID	Chr	Gene Location	Strand	Size (aa)	MW (kDa)	pI	Subcellular localization
GmSCPL1	GLYMA_02G158400	Chr02	17311794-17320940	+	74	8.89	6.54	Cy[a], Ch[a], Nu[b]
GmSCPL2	GLYMA_02G204700	Chr02	38993365-39000790	+	461	52.40	8.69	Pm[a], Ex[a], V[b],
GmSCPL3	GLYMA_03G070300	Chr03	16324083-16330477	+	462	52.92	5.75	Pm[a], Ex[ab], Ch[b]
GmSCPL4	GLYMA_03G090900	Chr03	26981330-26988564	+	230	26.55	4.89	Cy[a], Cy[b], Nu[b]
GmSCPL5	GLYMA_03G125200	Chr03	33883674-33887698	+	481	54.51	8.29	Ex[a], Pm[a], Ly[a],
GmSCPL6	GLYMA_03G125400	Chr03	33909174-33912239	+	462	51.19	5.13	Ex[ab], Ch[b]
GmSCPL7	GLYMA_03G125500	Chr03	33918983-33921895	+	459	51.36	5.49	Ex[ab], Ch[b]
GmSCPL8	GLYMA_03G125600	Chr03	33929465-33932783	+	461	51.60	6.37	Ex[ab], V[b]
GmSCPL9	GLYMA_04G155900	Chr04	36484956-36492419	-	483	54.80	5.73	Ly[a], ER[b], Cy[b]
GmSCPL10	GLYMA_04G200200	Chr04	47272565-47276234	+	469	52.65	6.72	Ex[ab], Ch[b]
GmSCPL11	GLYMA_04G240600	Chr04	50888492-50892907	+	473	52.90	6.07	Ly[a], V[b], Cy[b]
GmSCPL12	GLYMA_05G085800	Chr05	15943887-15945785	-	182	20.71	7.59	Pm[a], Nu[b], Ex[b]
GmSCPL13	GLYMA_06G047400	Chr06	3584404-3590095	-	385	43.66	6.27	Ly[a], Ex[a], V[b],
GmSCPL14	GLYMA_06G122800	Chr06	10009186-10013239	-	380	42.21	7.65	Ly[a], V[b], Ex[b]
GmSCPL15	GLYMA_06G165400	Chr06	13723574-13727195	-	470	52.77	6.72	Ex[ab], V[b]
GmSCPL16	GLYMA_06G181700	Chr06	15515149-15520025	+	417	47.35	5.82	Pm[ab], Ex[a], ER[b]
GmSCPL17	GLYMA_07G172400	Chr07	30030699-30035054	-	179	20.57	8.05	Ex[a], V[a], Nu[b],
GmSCPL18	GLYMA_07G192700	Chr07	36045428-36052232	+	486	54.73	5.42	Ex[a], V[a], Ly[a],
GmSCPL19	GLYMA_07G218400	Chr07	39141266-39142853	-	454	50.64	6.41	Pm[a], Ch[ab], Nu[b]
GmSCPL20	GLYMA_07G236600	Chr07	41820157-41824129	+	481	55.07	5.66	Ly[a], Ch[b], Ex[b]
GmSCPL21	GLYMA_08G008700	Chr08	682896-686087	-	464	52.58	6.76	Ex[a], Pm[a],
GmSCPL22	GLYMA_08G228700	Chr08	18660718-18661378	+	110	11.96	4.52	Nu[a], Ex[a], Cy[b]
GmSCPL23	GLYMA_08G245500	Chr08	21195233-21201784	-	471	53.35	9.14	Ly[a], V[b], Ex[b]
GmSCPL24	GLYMA_08G257800	Chr08	23162701-23166441	-	461	51.79	6.15	Ex[a], Ch[b], V[b]
GmSCPL25	GLYMA_09G049500	Chr09	4298942-4302333	-	497	54.65	4.81	V[a], Ch[b], Ex[b]
GmSCPL26	GLYMA_09G226700	Chr09	45147433-45152456	+	496	55.68	6.25	Ex[a], Ch[b], V[b]
GmSCPL27	GLYMA_09G249500	Chr09	47056885-47062577	+	506	56.71	6.18	V[a], Ly[a], Ch[b],
GmSCPL28	GLYMA_10G102100	Chr10	20236973-20240885	-	464	51.60	5.4	Ex[ab], Pm[a], V[b]
GmSCPL29	GLYMA_10G104300	Chr10	22636718-22663087	-	458	51.34	6.02	Cy[a], V[b], Ex[b]
GmSCPL30	GLYMA_11G099600	Chr11	7551710-7556030	-	473	53.44	8.67	Ly[a], Ex[b], ER[b]
GmSCPL31	GLYMA_11G099700	Chr11	7558753-7566979	-	466	52.87	6.24	Ly[a], Ex[ab], ER[b]
GmSCPL32	GLYMA_11G170600	Chr11	18267420-18268796	+	218	24.84	7.72	Ex[a], ER[b]
GmSCPL33	GLYMA_11G191200	Chr11	26434930-26442035	-	458	51.05	6.08	Cy[a], ER[b], Ex[b]
GmSCPL34	GLYMA_11G193800	Chr11	26725185-26729444	+	488	54.60	6.28	V[a], Ch[b], ER[b]
GmSCPL35	GLYMA_11G193900	Chr11	26730950-26734235	+	477	53.03	5.45	V[a], M[b], Nu[b]
GmSCPL36	GLYMA_11G207600	Chr11	29503252-29504278	-	187	21.22	5.7	Pm[a], Ex[a], Nu[b],
GmSCPL37	GLYMA_12G010100	Chr12	718029-722960	-	496	55.67	6.65	Ex[a], V[b], G[b]
GmSCPL38	GLYMA_12G025600	Chr12	1844961-1853018	-	482	54.42	6.89	Ly[a], V[b], ER[b]
GmSCPL39	GLYMA_12G025900	Chr12	1873092-1878613	-	472	53.41	8.32	Pm[a], Cy[b]
GmSCPL40	GLYMA_12G083100	Chr12	6586726-6592302	+	459	51.06	5.34	Cy[a], Ex[b], ER[b]
GmSCPL41	GLYMA_12G177900	Chr12	33699331-33703039	-	504	56.10	5.27	V[ab], Ch[b]
GmSCPL42	GLYMA_12G179400	Chr12	33972888-33973824	+	217	24.23	9.17	M[a], Ex[b], G[b]
GmSCPL43	GLYMA_13G028500	Chr13	8018267-8022397	+	490	54.82	5.96	Ly[a], V[b], Ex[b]
GmSCPL44	GLYMA_13G028800	Chr13	8182077-8185659	-	532	59.82	5.54	Pm[a], Ex[a], Ly[a],
GmSCPL45	GLYMA_13G029200	Chr13	8614070-8618166	+	488	54.20	6.14	Ex[ab], Pm[a], V[b]

GmSCPL46	GLYMA_13G075100	Chr13	17809545-17814049	+	93	10.17	5.17	Ex[ab], V[b]
GmSCPL47	GLYMA_13G075200	Chr13	17816382-17822985	+	293	34.27	7.07	Cy[a], Ex[a], Ch[b]
GmSCPL48	GLYMA_13G108900	Chr13	22294647-22296172	+	70	7.86	6.82	M[a], Cy[b]
GmSCPL49	GLYMA_13G183700	Chr13	29723379-29727554	-	493	55.52	5.73	V[a], Ex[a], Ly[a],
GmSCPL50	GLYMA_13G221900	Chr13	33488305-33492501	-	469	53.78	7.19	Ly[a], Nu[b], Cy[b]
GmSCPL51	GLYMA_13G243500	Chr13	35288591-35293220	-	478	54.05	6.33	Ly[a], V[a], Ex[a],
GmSCPL52	GLYMA_13G321200	Chr13	41546470-41550353	-	456	51.16	9.37	M[a], Pm[ab], ER[b],
GmSCPL53	GLYMA_13G321700	Chr13	41602238-41606724	+	458	51.07	4.84	Cy[a], ER[a], Ex[b],
GmSCPL54	GLYMA_13G322600	Chr13	41709896-41713941	+	506	56.48	5.49	V[ab], Ch[b]
GmSCPL55	GLYMA_14G080500	Chr14	6965564-6972041	+	498	55.49	5.45	Ly[a], Ex[a], ER[b],
GmSCPL56	GLYMA_14G122200	Chr14	17801426-17807621	-	456	50.73	6.5	Ex[a], Ly[a], ER[b]
GmSCPL57	GLYMA_15G070200	Chr15	5375186-5380088	+	482	54.39	5.96	V[a], Ex[a], Ly[a],
GmSCPL58	GLYMA_15G090300	Chr15	6952810-6957605	+	485	55.44	7.6	Pm[a], Cy[b], ER[b]
GmSCPL59	GLYMA_15G156200	Chr15	13086610-13091482	-	479	52.68	5.25	V[a], Ch[b], Cy[b]
GmSCPL60	GLYMA_16G082700	Chr16	9310673-9320591	-	498	55.69	5.76	Ly[a], V[b], Ch[b]
GmSCPL61	GLYMA_16G145300	Chr16	30604586-30608656	+	493	55.58	5.07	Pm[a], Ly[a], ER[b],
GmSCPL62	GLYMA_17G036900	Chr17	2694106-2696329	-	346	39.99	5.25	Pm[a], V[a], Nu[ab],
GmSCPL63	GLYMA_17G037000	Chr17	2697476-2701132	-	482	54.83	5.68	Ly[a], Ex[ab], Ch[b]
GmSCPL64	GLYMA_17G073200	Chr17	5720584-5728156	-	460	51.99	8.68	Ly[a], Pm[a], Ex[ab],
GmSCPL65	GLYMA_17G244900	Chr17	40030818-40036322	-	470	52.27	5.65	Ex[a], Ly[a], ER[b],
GmSCPL66	GLYMA_18G098200	Chr18	10197299-10198615	+	134	15.19	8.82	Nu[ab], M[ab]
GmSCPL67	GLYMA_18G202200	Chr18	48297961-48298505	+	94	10.34	4.61	Ex[a], Nu[b], Ch[b]
GmSCPL68	GLYMA_18G242900	Chr18	53113341-53119592	-	506	56.68	6.07	V[a], Ly[a], Ch[b],
GmSCPL69	GLYMA_18G266700	Chr18	55105365-55112399	-	467	52.78	8.51	Ly[a], Ch[b], V[b]
GmSCPL70	GLYMA_18G282200	Chr18	56302289-56306074	-	461	51.68	6.02	Ex[ab], Ch[b]
GmSCPL71	GLYMA_19G128000	Chr19	38717520-38722550	+	374	42.34	8	Ex[ab], Ly[a], V[b]
GmSCPL72	GLYMA_19G128100	Chr19	38728037-38732422	+	462	51.38	6.19	Pm[a], Ex[a], V[b],
GmSCPL73	GLYMA_19G128200	Chr19	38739769-38743404	+	460	51.18	6.42	Ex[a], Ch[b], Cy[b]
GmSCPL74	GLYMA_20G014900	Chr20	1375276-1376799	+	300	33.38	6.5	Pm[a], Ex[a], M[b],
GmSCPL75	GLYMA_20G015100	Chr20	1379240-1380596	-	445	49.87	6.57	Pm[a], Ch[a], V[b],
GmSCPL76	GLYMA_20G015300	Chr20	1391806-1392346	-	161	18.63	4.9	Cy[ab], Ch[b]
GmSCPL77	GLYMA_20G015400	Chr20	1393894-1395216	-	440	49.38	6.42	Ch[a], Pm[ab], ER[b]
GmSCPL78	GLYMA_20G015600	Chr20	1406934-1408603	-	438	49.15	9.15	M[a], Ch[ab], V[b]
GmSCPL79	GLYMA_20G016800	Chr20	1545314-1546525	-	393	43.97	6.66	Ch[ab], M[ab]
GmSCPL80	GLYMA_20G043400	Chr20	7902915-7918077	+	373	43.12	5.6	Ly[a], Ex[a], Ch[b],
GmSCPL81	GLYMA_20G056200	Chr20	13855383-13860273	-	483	53.92	5.34	Ly[a], Ex[a], V[b],
GmSCPL82	GLYMA_20G071400	Chr20	25407177-25407413	+	78	8.91	4.23	Ex[a], Cy[b], Ch[b]
GmSCPL83	GLYMA_20G178600	Chr20	41603247-41609549	+	460	52.04	8.72	Ly[a], V[b], G[b]
GmSCPL84	GLYMA_U001900	Super ContigKZ847665	39-938	-	299	33.08	8.86	Pm[a], Ch[ab], M[b]

MW- Molecular Weight, pI- Isoelectric point, GRAVY- Grand average of hydropathy, Pm- Plasma membrane, Ch-Chloroplast, Nu-Nuclear, Ex-Extracellular, V-Vacuole, Ly-Lysosomal, M-Mitochondrial, Cy- Cytoplasm, and ER- Endoplasmic reticulum. [a]subcellular localization according to CELLO, [b]subcellular localization according to WoLF PSORT.

play in secondary metabolism, as seen in *Arabidopsis* and rice (Fraser et al. 2007). Group II in soybean, containing the largest number of genes, showed notable clustering with *SCPL* genes from rice and *Arabidopsis*. Hence, these results suggest that considerable diversity exists for the *GmSCPLs* in soybean.

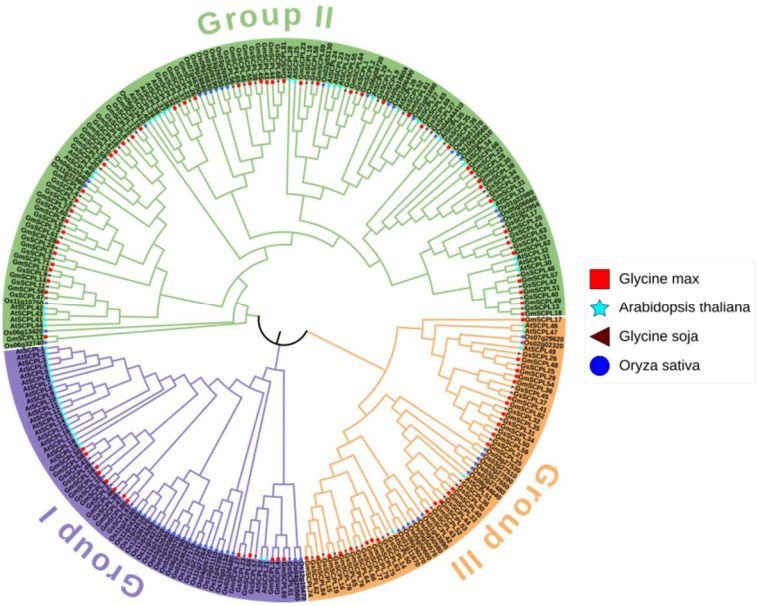

Fig. 1. Phylogenetic relationships of SCPL proteins in soybean, rice, and *Arabidopsis* based on amino acid sequence alignment. The complete amino acid sequences were aligned using MUSCLE algorithm and a Maximum-likelihood method with MEGA X. *SCPL* genes from different groups are marked with different colors: Group I (violet), Group II (green), and Group III (orange). Different species were marked with different colored shapes: *Glycine max* (red squares), *Glycine soja* (brown left pointing triangle), rice (blue circles), and *Arabidopsis* (teal stars).

As shown in Fig. 2a, the organization of the introns and exons of *GmSCPLs* was variable. One to fourteen exons were found in *GmSCPLs*. The maximum number of exons (14) was found in *GmSCPL60*, *GmSCPL27*, *GmSCPL68*, *GmSCPL3*, *GmSCPL50*, and *GmSCPL58* (Fig. 2a). Approximately 26% of *GmSCPL* genes contain 10 exons, 17% have 9 exons, 11% have 8 exons, and 7% contain 14 exons. The remaining 39% of genes have between 1 to 7 or 11 to 14 exons. The distribution of exons and introns is complex, with a different structural pattern of exon/intron composition even within the same phylogenetic group. Using the MEME web server, 10 motifs were identified and the length of the motifs ranged from 15 to 29 amino acids (Fig. 2b). Most of the GmSCPL proteins shared seven common motifs (motifs 1-5, 7 and 8) (Fig. 2b). SCPL proteins within the same phylogenetic group generally exhibited identical or highly similar motifs. For instance, most proteins in Groups I and II contained conserved motifs 8, 3, 4,

1, 7, 9, and 2. However, motif 9 was completely absent in Group III (Fig. 2b). Among all SCPL genes in soybean, GmSCPL44 had the longest sequence and included 9 motifs (motifs 1 and 3-10) (Fig. 2b). This extended motif composition in *GmSCPL44* suggests that the gene may have diverged in function from other members of the *SCPL* family. Domain analysis showed that all GmSCPL proteins had the conserved Peptidase_S10 domain, which is necessary for enzymatic activity. This domain's conservation across poplar, grape, and wheat species emphasizes its importance in SCPL protein catalysis (Zhu et al. 2018, Xu 2021 and Wang et al. 2022).

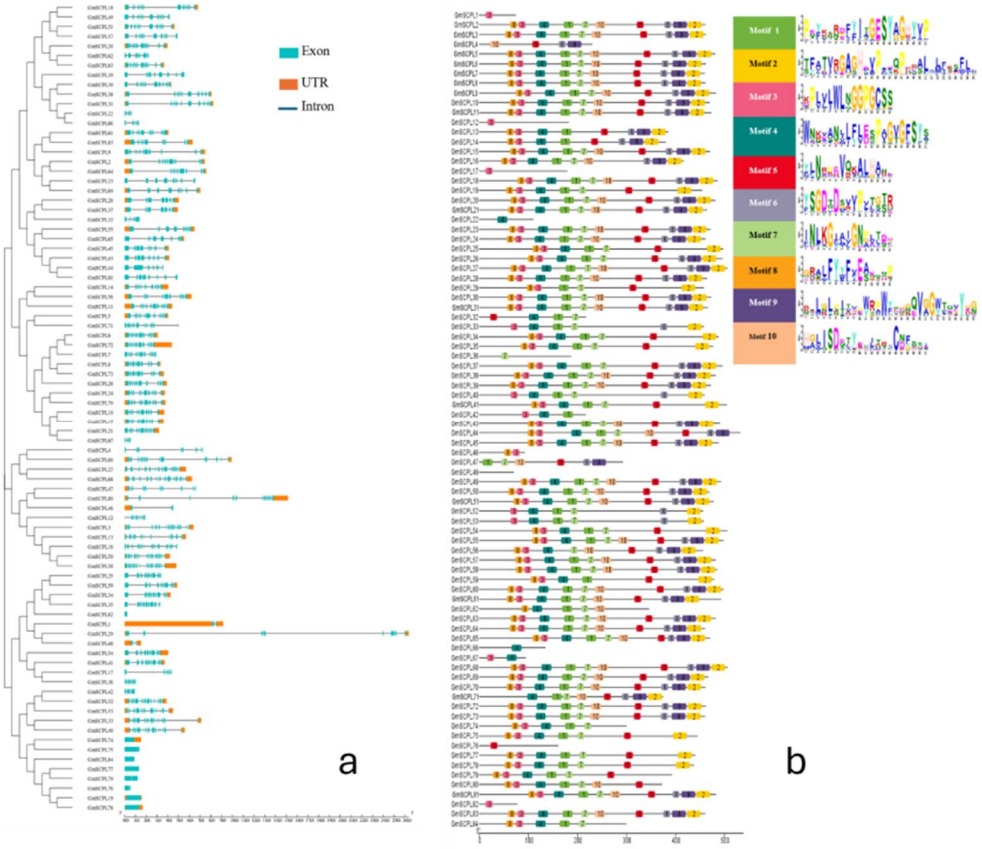

Fig. 2. Gene structure and conserved motifs analysis of *GmSCPL* genes. (a) Gene structure of *GmSCPLs*. The exons, UTRs, and introns are marked with teal-colored boxes, orange-colored boxes, and single lines, respectively. (b) Conserved motif compositions of GmSCPL proteins were identified using MEME. Each color represents a specific motif.

The distribution of all identified 84 *GmSCPL* genes spanned across 19 chromosomes, except for chromosome 1 (Chr01) (Fig. 3). A total of 12 *GmSCPL* genes were clustered on Chr 13, making it the chromosome with the highest density of *GmSCPL* genes, followed

by Chr 20, which contained 10 *GmSCPL* genes (Fig. 3). A total of 60 *GmSCPL* genes were distributed throughout sixteen chromosomes (2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, and 19) (Fig. 3). Each chromosome contains between two and seven *GmSCPL* genes. Chr 5 harbored only one *GmSCPL* gene, indicating an uneven distribution of the 83 *GmSCPL* genes across 20 chromosomes of soybean (Fig. 3).

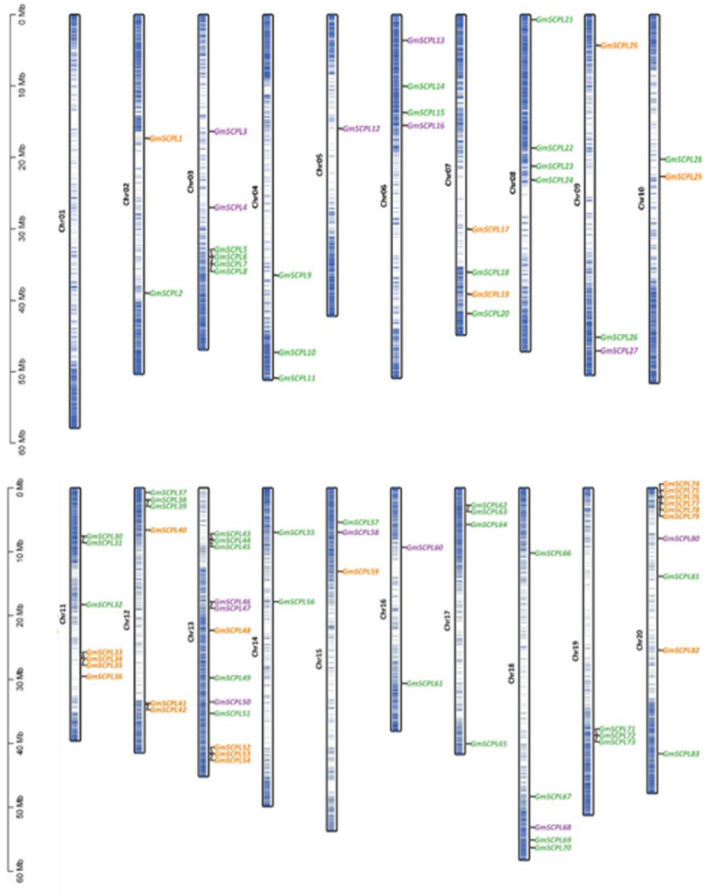

Fig. 3. Chromosomal distribution of *SCPL* genes on different chromosomes of soybean. The chromosome number (Chr01-Chr20) is indicated at the left of each chromosome. The scale is in megabase pairs (Mbp).

In this study, the similarity percentage of 84 *GmSCPL* genes was calculated based on their coding nucleotide sequences and identified a total of 31 duplication events. These included 27 gene pairs associated with whole-genome duplication (WGD) and 4 pairs with tandem duplication (TD) types (Fig. 4a). These duplicated gene pairs were distributed across 18 of the 20 chromosomes, with no duplication events found on

chromosomes 1 and 5. Chromosome 13 contained the highest number of duplication events (7), while chromosomes 2 and 4 had the fewest, each with just one duplication event (Fig. 4a). To further explore the evolutionary connections within the *SCPL* gene family, we performed a syntenic analysis to identify orthologous *SCPL* gene pairs between soybean and both monocotyledonous plants (*O. sativa*) and dicotyledonous plants (*A. thaliana*, *G. soja*, and *V. vinifera*) using the Multiple Collinearity Scan Toolkit (MCScanX). This analysis revealed 33 orthologous gene pairs between *G. max* and *A. thaliana* (*G. max*-A. *thaliana*), 117 pairs between *G. max* and *G. soja* (*G. max*-*G. soja*), 36 pairs between *G. max* and *V. vinifera* (*G. max*-*V. vinifera*), and only eight orthologous pairs between *G. max* and *O. sativa* (*G. max*-*O. sativa*) (Fig. 4b). These results suggest a strong evolutionary link between the *SCPL* genes in soybean and *G. soja*, likely due to their close genetic relationship.

The rates of synonymous (Ks) and non-synonymous (Ka) substitutions were calculated for all duplicated gene pairs in *Glycine max* to assess the selection pressure driving functional diversification. Values of Ka/Ks greater than 1, equal to 1, and less than 1 indicate positive selection, neutral selection, and purifying selection, respectively. A total of 31 duplicated gene pairs were identified, with 27 showing segmental duplication and 4 showing tandem duplication (Table 2). Notably, all 31 duplicated gene pairs were found to be undergoing purifying selection. To further explore the evolutionary forces acting on the 84 *GmSCPL* genes, Ka/Ks ratios were calculated for each duplicated gene pair. All Ka/Ks ratios were below 0.8, ranging from 0.11 to 0.71, with an average ratio of 0.34. Specifically, 26% of the duplicated gene pairs had Ka/Ks ratios between 0.2 to 0.3 and 0.3 to 0.4, 16% ranged from 0.1 to 0.2, and 23% fell within the 0.4 to 0.5 range (Table 2). The Ka/Ks ratios for the 4 tandem duplicated gene pairs were between 0.400 and 0.442 (Table 2).

GmSCPL genes were grouped into functional classes using GO enrichment analysis (ShinyGO 0.80) to determine their possible roles. The three GO categories of biological processes (BP), molecular function (MF), and cellular components (CC) was significantly enriched, with an FDR-adjusted $p < 0.05$. The majority of genes were associated with biological processes like proteolysis (GO:0006508), proteolysis involved in the cellular protein catabolic process (GO:0051603), cellular protein catabolic process (GO:0044257), protein catabolic process (GO:0030163), cellular macromolecule catabolic process (GO:0044265), organonitrogen compound catabolic process (GO:1901565), macromolecule catabolic process (GO:0009057), and cellular catabolic process (GO:0044248) (Fig. 5a). Many genes were linked to serine-type carboxypeptidase, exopeptidase, peptidase, hydrolase, and peptidase activity in the molecular function category (Fig. 5b). Most genes were related with extracellular region (GO:0005576), vacuole (GO:0005773), and space (GO:0005615) in the cellular components category (Fig. 5c). These GO terms suggest that the *GmSCPL* genes may have a role in the stress response and the regulation of plant cell wall dynamics. To further explore the interactions among GmSCPLs proteins, we constructed protein-protein interaction

networks of all the identified *GmSCPLs* genes using STRING online software. The possible network obtained from STRING was subsequently visualized in the network analyst platform. The sub-network 1 comprised 84 nodes (proteins) and 56 edges (interactions) with the PPI enrichment p-value< 1.0e-16 at the medium confidence parameter level. From PPI network analysis, we identified three hub genes viz., K7N217_SOYBN (*GmSCPL82*), K7LX88_SOYBN (*GmSCPL46*), K7KVS1_SOYBN (*GmSCPL16*) that may be associated with soybean abiotic stress tolerance (Fig. 5d).

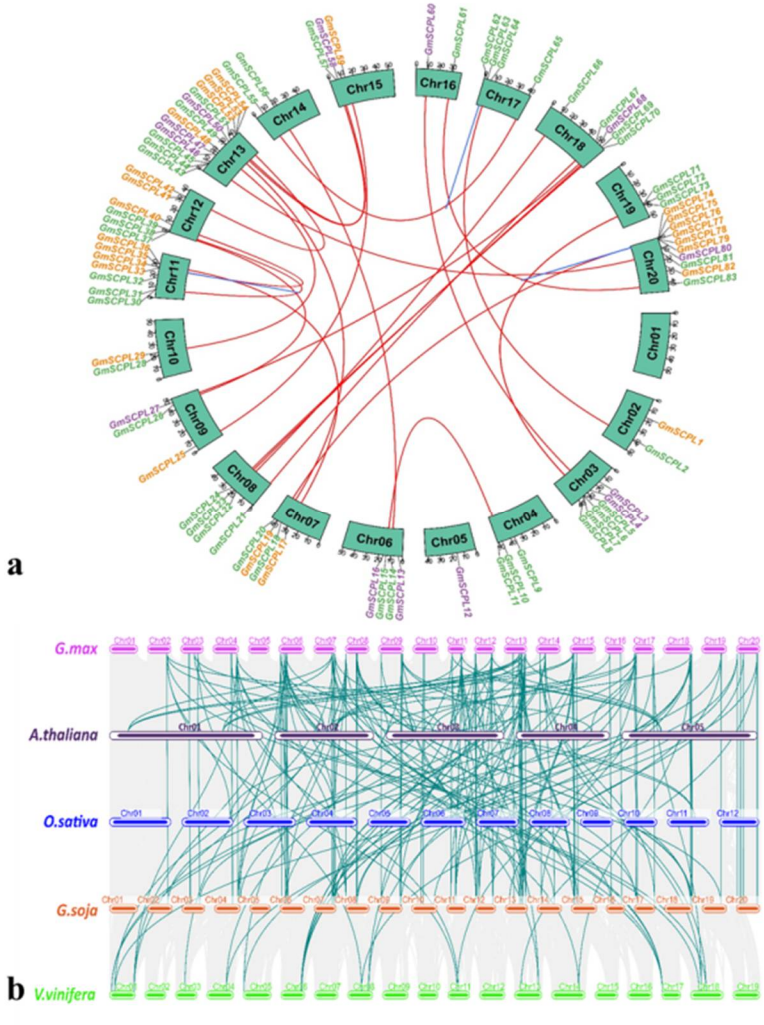

Fig. 4. Gene duplication and synteny analysis of *GmSCPLs*. (a) WGD/segmental and tandem duplications of *GmSCPLs* were mapped on the soybean genome using a Circos plot. Red lines indicate segmental duplications and blue lines indicate tandem duplications. (b) Syntenic relationships of *SCPLs* between *G. max* and four other species (*A. thaliana*, *O. sativa*, *G. soja*, and *V. vinifera*) were shown in teal color

Table 2. Duplicated *SCPL* genes and estimated duplication events in soybean.

Seq_1	Seq_2	Ka	Ks	Ka/Ks	Type of Mutation/ Evolution	Duplication
GmSCPL24	*GmSCPL70*	0.012435518	0.111223801	0.111806268	Purifying	Segmental
GmSCPL18	*GmSCPL49*	0.013436203	0.106774865	0.125836761	Purifying	Segmental
GmSCPL14	*GmSCPL56*	0.067924158	0.493928675	0.13751815	Purifying	Segmental
GmSCPL51	*GmSCPL57*	0.02111361	0.125325078	0.168470753	Purifying	Segmental
GmSCPL10	*GmSCPL15*	0.022696767	0.116550627	0.194737412	Purifying	Segmental
GmSCPL23	*GmSCPL69*	0.023309312	0.114801355	0.203040392	Purifying	Segmental
GmSCPL8	*GmSCPL73*	0.027249186	0.134198032	0.203052053	Purifying	Segmental
GmSCPL33	*GmSCPL40*	0.02882199	0.132440423	0.217622306	Purifying	Segmental
GmSCPL27	*GmSCPL68*	0.031459396	0.120817003	0.26038881	Purifying	Segmental
GmSCPL19	*GmSCPL78*	0.066733637	0.240436208	0.277552359	Purifying	Segmental
GmSCPL2	*GmSCPL64*	0.02413724	0.083913327	0.287644893	Purifying	Segmental
GmSCPL47	*GmSCPL80*	0.351416552	1.184936605	0.296569918	Purifying	Segmental
GmSCPL44	*GmSCPL81*	0.082992211	0.278719602	0.297762376	Purifying	Segmental
GmSCPL55	*GmSCPL65*	0.025651645	0.084907322	0.302113466	Purifying	Segmental
GmSCPL26	*GmSCPL37*	0.019127733	0.061060745	0.313257436	Purifying	Segmental
GmSCPL25	*GmSCPL59*	0.049677808	0.151704257	0.327464826	Purifying	Segmental
GmSCPL61	*GmSCPL83*	0.152773574	0.445225755	0.343137324	Purifying	Segmental
GmSCPL6	*GmSCPL72*	0.040117168	0.111998992	0.358192226	Purifying	Segmental
GmSCPL42	*GmSCPL52*	0.078023046	0.215138372	0.362664482	Purifying	Segmental
GmSCPL30	*GmSCPL38*	0.053792539	0.138630337	0.388028622	Purifying	Segmental
GmSCPL21	*GmSCPL67*	0.044402607	0.112058867	0.396243586	Purifying	Segmental
GmSCPL34	*GmSCPL35*	0.099995118	0.249408626	0.400928868	Purifying	Tandem
GmSCPL74	*GmSCPL75*	0.055603064	0.127345207	0.436632562	Purifying	Tandem
GmSCPL77	*GmSCPL79*	0.016702927	0.037850984	0.441281176	Purifying	Tandem
GmSCPL62	*GmSCPL63*	0.059980141	0.135665536	0.442117745	Purifying	Tandem
GmSCPL29	*GmSCPL48*	0.063597402	0.143291428	0.44383257	Purifying	Segmental
GmSCPL50	*GmSCPL58*	0.059937376	0.1295813	0.462546495	Purifying	Segmental
GmSCPL5	*GmSCPL71*	0.071883702	0.148981239	0.482501707	Purifying	Segmental
GmSCPL4	*GmSCPL60*	0.092500911	0.165046414	0.560453926	Purifying	Segmental
GmSCPL17	*GmSCPL36*	0.246653518	0.402261788	0.613166662	Purifying	Segmental
GmSCPL22	*GmSCPL66*	0.195370455	0.272680625	0.716480883	Purifying	Segmental

The upstream promoter regions of genes contain various CAREs that modulate gene expression. In this study, 2 kb promoter regions (upstream of the start codon) of all *GmSCPL* genes were investigated. The results revealed that eighteen types of CAREs were identified in the promoters of *GmSCPL* genes (Fig. 6a). A total of 1,631 CAREs were found across all *GmSCPL* genes. Most of the identified CAREs were light-responsive elements, accounting for 44%, followed by hormone-responsive elements at 30%, environmental stress-responsive elements at 20%, and plant growth-related elements at

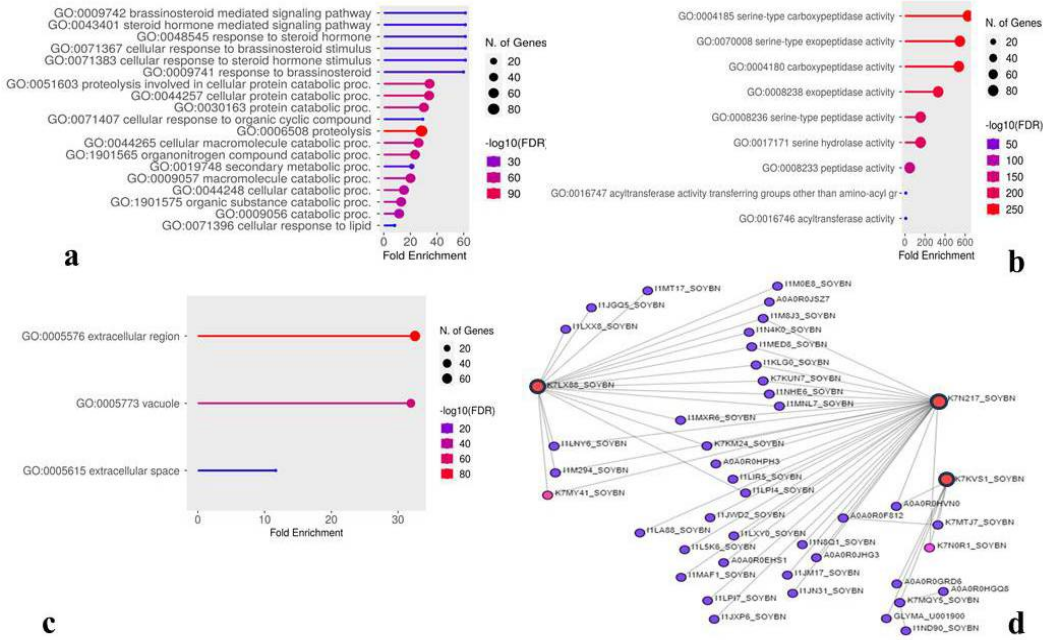

Fig. 5. Gene ontology enrichment and protein-protein interaction network analyses of *GmSCPLs*. (a) Biological Processes, (b) Molecular Functions, and (c) Cellular Components. Red plots represent higher gene expression for specific functions, while smaller blue plots indicate lower expression. (d) Graphical depiction of the selective hub module of the sub-network. Hub genes are highlighted in red. Edges of the hub are shown in gray.

6% (Fig. 6b). Consistent with our findings, similar *cis*-acting elements were identified in the promoter regions of *SCPL* genes in poplar (Zhu et al. 2018). To gain insights into the potential functions of *GmSCPLs*, we analyzed RNA sequencing data to examine the transcript levels of all members under heat, salinity, and drought stress conditions, assessing their expression profiles across various abiotic stresses. Heat stress induced the most dramatic up-regulation across the gene family. The genes most significantly up-regulated under heat stress were *GmSCPL20*, *GmSCPL33*, and *GmSCPL76*; under drought stress, *GmSCPL40*, *GmSCPL64*, and *GmSCPL69*; and salt stress, *GmSCPL2*, *GmSCPL27*, and *GmSCPL37* (Fig. 7). *GmSCPL2* emerged as the only gene consistently up-regulated across all three stress conditions, indicating a potential central role in general abiotic stress responses. Simultaneously, *GmSCPL3*, *GmSCPL5*, *GmSCPL10*, *GmSCPL15*, *GmSCPL19*, *GmSCPL24*, and *GmSCPL25* displayed increased expression under drought and heat stress but were down-regulated in response to salt stress. In contrast, *GmSCPL26* and *GmSCPL27* were up-regulated in heat and salt stress conditions but exhibited lower expression levels under drought stress (Fig. 7). Extensive research has explored the roles of *SCPLs* in coping with different abiotic stresses. For example, *VvSCPL* genes of grape have responded to drought- and waterlogging-stress treatments,

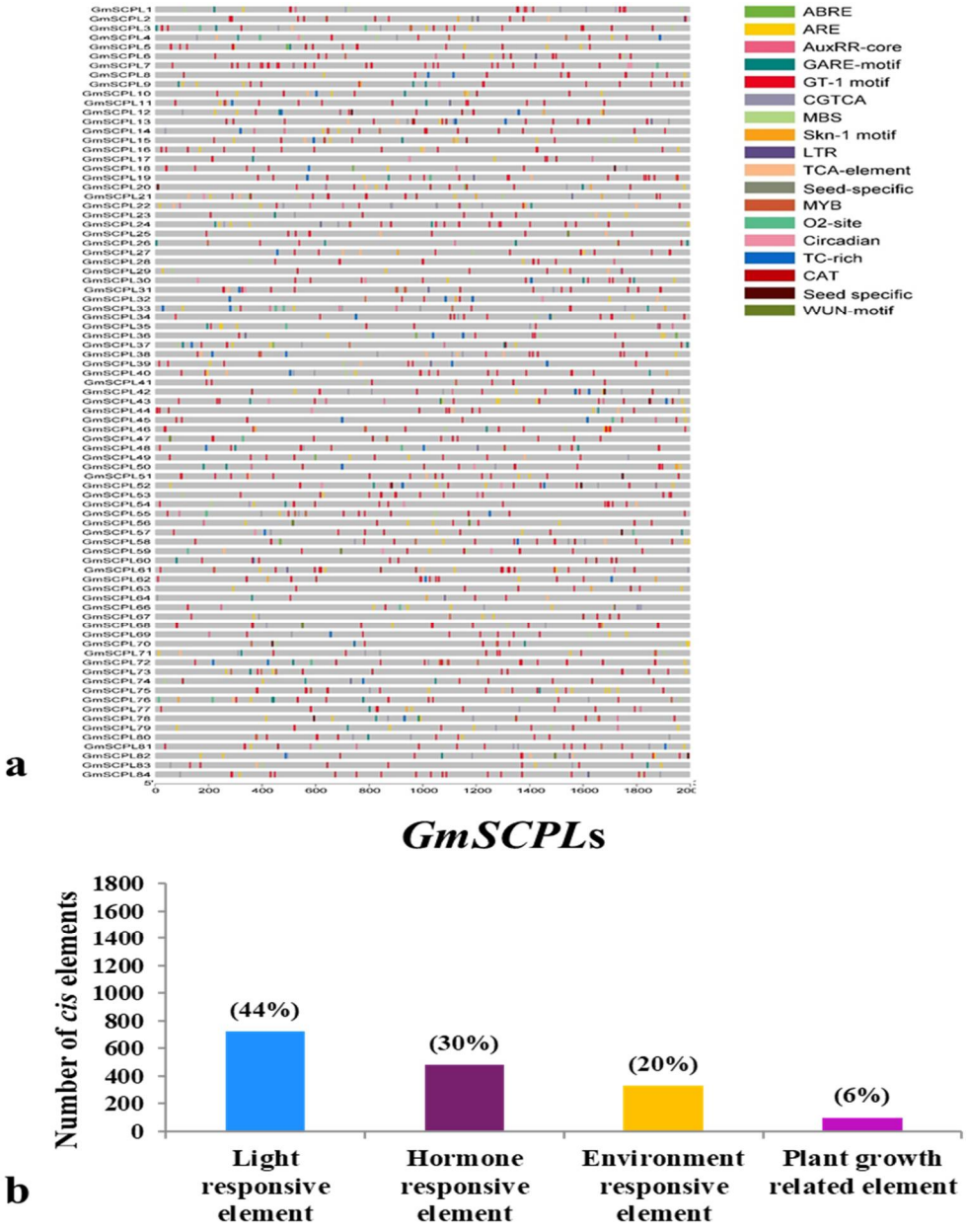

Fig. 6. Analysis of c*is*-acting regulatory elements in *GmSCPLs*. (a) Diagram showing the approximate locations of predicted *cis*-regulatory elements in each *GmSCPL* gene, as identified by the PlantCare database. Different colored boxes indicate various promoter elements and their respective positions. (b) The number of *cis*-regulatory elements related to light responsive elements, environmental stress-responsive elements, hormone-responsive elements, and plant growth-related elements in *GmSCPLs*.

Fig. 7. Expression profiling of *SCPL* genes under drought, heat, and salt stresses in soybean. Hierarchical clustering displays *SCPL* gene expression levels, with color scaling to indicate regulation: red represents up-regulation, and green represents down-regulation across stress conditions.

which indicated their roles in abiotic stress responses (Wang et al. 2021). Rice *SCPL* gene, such as *OsBISCPL1*, contributes to the development of resistance against oxidative stress (Liu et al. 2008). Our research offers a comprehensive analysis of the SCPL gene family, providing valuable insights into the evolutionary mechanisms of soybean SCPL genes. This study paves the way for future functional analyses and breeding strategies aimed at enhancing stress tolerance in soybean.

Acknowledgment

We would like to express our sincere gratitude to the Jagannath University and University Grants Commission, Dhaka, Bangladesh for their generous financial support in conducting this research.

References

Afgan E, Baker D, Batut B, Van Den Beek M, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Grüning BA and **Guerler A** (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. **46**(W1): W537-W544.

Agarwal V, Tikhonov A, Metlitskaya A, Severinov K and **Nair SK** (2012) Structure and function of a serine carboxypeptidase adapted for degradation of the protein synthesis antibiotic microcin C7. Proc. Natl. Acad. Sci. USA. **109**(12): 4425-4430.

Ahmad MZ, Li P, She G, Xia E, Benedito VA, Wan XC and **Zhao J** (2020) Genome-wide analysis of serine carboxypeptidase-like acyltransferase gene family for evolution and characterization of enzymes involved in the biosynthesis of galloylated catechins in the tea plant (*Camellia sinensis*). Front. Plant Sci. **11**: 848.

Bailey TL, Johnson J, Grant CE and **Noble WS** (2015) The MEME suite. Nucleic Acids Res. **43**(W1): W39-W49.

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y and **Xia R** (2020) TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant. **13**(8): 1194-1202.

Chen J, Li WQ and **Jia YX** (2020) The serine carboxypeptidase-like gene *SCPL41* negatively regulates membrane lipid metabolism in *Arabidopsis thaliana*. Plants **9**(6): 696.

Ciarkowska A, Ostrowski M and **Jakubowska A** (2016) Abiotic stress and phytohormones affect enzymic activity of 1-O-(indole-3-acetyl)-β-D-glucose: myo-inositol indoleacetyl transferase from rice (*Oryza sativa*). J. Plant Physiol. **205**: 93-96.

Cohen I, Zandalinas SI, Fritschi FB, Sengupta S, Fichman Y, Azad RK and **Mittler R** (2021) The impact of water deficit and heat stress combination on the molecular response, physiology, and seed production of soybean. Physiol. Plant **172**(1): 41-52.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**(4): 783-791.

Feng Y and **Xue Q** (2006) The serine carboxypeptidase-like gene family of rice (*Oryza sativa* L. ssp. japonica). Funct. Integr. Genomics **6**: 14-24.

Fraser CM, Rider LW and **Chapple C** (2005) An expression and bioinformatics analysis of the *Arabidopsis* serine carboxypeptidase-like gene family. Plant Physiol. **138**(2): 1136-1148.

Fraser CM, Thompson MG, Shirley AM, Ralph J, Schoenherr JA, Sinlapadech T, Hall MC and **Chapple C** (2007) Related *Arabidopsis* serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. Plant Physiol. **144**(4): 1986-1999.

Gasteiger E (2005) Protein identification and analysis tools on the ExPASy server. *In*: Walker JM (Ed.), The Proteomics Protocols Handbook, Humana Press Inc., pp. 571-607.

Hartman GL, West ED and **Herman TK** (2011) Crops that feed the world 2. Soybean-worldwide production, use, and constraints caused by pathogens and pests. Food Secur. **3**: 5-17.

He L, Liu Q, and **Han S** (2024) Genome-Wide Analysis of Serine Carboxypeptidase-like Genes in Soybean and Their Roles in Stress Resistance. Int. J. Mol. Sci. **25**(12): 6712.

Horton P, Park KJ, Obayashi T, Fujita N, Harada H, Adams-Collier CJ and **Nakai K** (2007) WoLF PSORT: protein localization predictor. Nucleic Acids Res. **35**(suppl_2): W585-W587.

Hu B, Jin J, Guo AY, Zhang H, Luo J and **Gao G** (2015) GSDS 2.0: an upgraded gene feature visualization server. Bioinformatics **31**(8): 1296-1297.

Hurst LD (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet. **18**(9): 486-487.

Jiang P, Gao J, Mu J, Duan L, Gu Y, Han S, Chen L, Li Y, Yan Y and **Li X** (2020) Interaction between serine carboxypeptidase-like protein TtGS5 and Annexin D1 in developing seeds of *Triticum timopheevi*. J. Appl. Genet. **61**: 151-162.

Kumar S, Stecher G, Li M, Knyaz C and **Tamura K** (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. **35**(6): 1547-1549.

Lescot M, Déhais P, Thijs G, Marchal K, Moreau Y, Van de Peer Y, Rouzé P and **Rombauts S** (2002) PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for *in silico* analysis of promoter sequences. Nucleic Acids Res. **30**(1): 325-327.

Letunic I and **Bork P** (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. **49**(W1): W293-W296.

Liu H, Wang X, Zhang H, Yang Y, Ge X and **Song F** (2008) A rice serine carboxypeptidase-like gene *OsBISCPL1* is involved in regulation of defense responses against biotic and oxidative stress. Gene. **420**(1): 57-65.

Lynch M and **Conery JS** (2000) The evolutionary fate and consequences of duplicate genes. Sci. **290**(5494): 1151-1155.

Liu Y, Ce F, Tang H, Tian G, Yang L, Qian W and **Dong H** (2022) Genome-wide analysis of the serine carboxypeptidase-like (SCPL) proteins in *Brassica napus* L. Plant Physiol. Biochem. **186**: 310-321.

Mazhar HS, Shafiq M, Ali H, Ashfaq M, Anwar A, Tabassum J, Ali Q, Jilani G, Awais M, Sahu R and **Javed MA** (2023) Genome-wide identification, and *in silico* expression analysis of YABBY gene family in response to biotic and abiotic stresses in potato (*Solanum tuberosum*). Genes. **14**(4): 824.

Milkowski C and **Strack D** (2004) Serine carboxypeptidase-like acyltransferases. Phytochemistry **65**(5): 517-524.

Miyake Y, Tanaka K, Okubo H, Sasaki S, Furukawa S and **Arakawa M** (2018) Soy isoflavone intake and prevalence of depressive symptoms during pregnancy in Japan: baseline data from the Kyushu Okinawa Maternal and Child Health Study. Eur. J. Nutr. **57**: 441-450.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R and **Finn RD** (2018) HMMER web server: 2018 update. Nucleic Acids Res. **46**(W1): W200-W204.

Qin P, Wang T and **Luo Y** (2022) A review on plant-based proteins from soybean: Health benefits and soy product development. J. Agric. Food Res. **7**: 100265.

Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P and **Jensen LJ** (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. **47**(D1): D607-D613.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M and **Kumar S** (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**(10): 2731-2739.

Wang L, Guo K, Li Y, Tu Y, Hu H, Wang B, Cui X and **Peng L** (2010) Expression profiling and integrative analysis of the *CESA/CSL* superfamily in rice. BMC Plant Biol. **10**: 1-6.

Wang XC, Wu WM, Zhou BB, Wang ZW, Qian YM, Bo WA and **Yan LC** (2021) Genome-wide analysis of the SCPL gene family in grape (*Vitis vinifera* L.). J. Integr. Agric. **20**(10): 2666-2679.

Wang Y, Zhao J, Deng X, Wang P, Geng S, Gao W, Guo P, Chen Q, Li C and **Qu Y** (2022) Genome-wide analysis of serine carboxypeptidase-like protein (SCPL) family and functional validation of *Gh_SCPL42* unchromosome conferring cotton Verticillium wilt stress in *Gossypium hirsutum*. BMC Plant Biol. **22**(1): 421.

Wen J, Li J and **Walker JC** (2012) Overexpression of a serine carboxypeptidase increases carpel number and seed production in *Arabidopsis thaliana*. Food Energy Secur. **1**(1): 61-69.

Yu CS, Chen YC, Lu CH and **Hwang JK** (2006) Prediction of protein subcellular localization. Proteins. **64**(3): 643-651.

Zhou G, Soufan O, Ewald J, Hancock RE, Basu N and **Xia J** (2019) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. **47**(W1): W234-W241.

Xu X, Zhang L, Zhao W, Fu L, Han Y, Wang K, Yan L, Li Y, Zhang XH and **Min DH** (2021) Genome-wide analysis of the serine carboxypeptidase-like protein family in *Triticum aestivum* reveals *TaSCPL184-6D* is involved in abiotic stress response. BMC Genomics **22**(1): 350.

Zhu D, Chu W, Wang Y, Yan H, Chen Z, Xiang Y (2018) Genome-wide identification, classification and expression analysis of the serine carboxypeptidase-like protein family in poplar. Physiol. Plant **162**(3): 333-352.

(Manuscript received on 16 November, 2024; revised on 9 December, 2024)