



OPEN

Comprehensive characterization of protease inhibiting gene family, *cis*-regulatory elements, and protein interaction network in linseed and their expression upon bud fly infestation

Chandra Mohan Singh¹, Bhupendra Kumar Singh^{2✉}, Shalini Purwar³, Beena Nair⁴, Ruchi¹, Amar Patel⁵, Saurabh Singh⁵ & Vikender Kaur^{6✉}

Linseed, also known as flax is an important oilseed crop with many potential uses in paint, textile, food and pharmaceutical industries. Susceptibility to bud fly (*Dasyneura lini* Barnes) infestation is a serious biotic concern leading to severe yield penalty in linseed. Protease inhibitors (PIs) are potential candidates that activate during the insect-pest attack and modulate the resistance. In the present study, we explored the PI candidates in the linseed genome and a total of 100 LuPI genes were identified and grouped into five distinct subgroups. The analysis of *cis*-acting elements revealed that almost all LuPI promoters contain several regulatory elementary related to growth and development, hormonal regulation and stress responses. Across the subfamilies of PIs, the specific domains are consistently found conserved in all protein sequences. The tissue-specific in-silico expression pattern via RNA-seq revealed that all the genes were regulated during different stress. The expression through qRT-PCR of 15 genes revealed the significant up-regulation of *LuPI-24*, *LuPI-40*, *LuPI-49*, *LuPI-53*, and *LuPI-63* upon bud fly infestation in resistant genotype EC0099001 and resistant check variety Neela. This study establishes a foundation resource for comprehending the structural, functional, and evolutionary dimensions of protease inhibitors in linseed.

Keywords Flax, *Linum*, Protease inhibitors, Bud fly, Gene expression, Phylogenetic analysis

Abbreviations

ABRE	ABA response elements
ARE	Anaerobic induction elements
BBI	Bowman-birk inhibitor
CDD	Conserved domain database
CPI	Cysteine proteinase inhibitors
ERE	Ethylene responsive elements
Gly	Glycine
GSDS	Gene structure display server
GO	Gene ontology

¹Department of Genetics and Plant Breeding, Banda University of Agriculture and Technology, Banda, Uttar Pradesh 210 001, India. ²Department of Entomology, Banda University of Agriculture and Technology, Banda, Uttar Pradesh 210 001, India. ³Department of Basic and Social Sciences, Banda University of Agriculture and Technology, Banda, Uttar Pradesh 210 001, India. ⁴AICRP on Linseed and Mustard, College of Agriculture, Dr. PDKV-Akola, Nagpur, Maharashtra 440 001, India. ⁵AICRP on Linseed and Sesame, Oilseed Research Station, Banda University of Agriculture and Technology, Mauranipur, Uttar Pradesh 282 204, India. ⁶Division of Germplasm Evaluation, Indian Council of Agricultural Research-National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi 110 012, India. ✉email: bhupendrakumarsingh@gmail.com; kaurvikender@gmail.com

KPI	Kunitz proteinase inhibitors
ML	Maximum likelihood
PCA	Principal component analysis
PI	Protease inhibitor
SRE	Stress responsiveness elements
Tryp_alpha_amyl	Trypsin alpha amylase inhibitor
UPI	Unusual serine protease inhibitor

Linseed (*Linum usitatissimum* L.) is an important multipurpose industrial crop used in the textile and paint industries, nutraceuticals, pharmaceutical and healthcare sectors, etc.^{1,2} It is a potential crop and can be cultivated in a wider range of ecologies such as rainfed, irrigated and high-temperature regimes. Canada is the leading producer and exporter of flax (~ 23% of world production) followed by Kazakhstan and Russia³. Presently, linseed occupies an area of 32.23 lakh ha yielding 30.68 lakh tonnes with an average productivity of 952 kg/ha in the world, whereas in India, it occupies an area of 1.7 lakh ha with a production and productivity of about 1 lakh tonnes and 574 kg/ha, respectively⁴. Various biotic and abiotic stresses adversely affect the linseed production and productivity^{5–8}. Among the biotic stresses, linseed bud fly (*Dasyneura lini* Barnes) is a major production constraint that leads to severe yield penalty of up to 90%⁹. Its infestation occurs during the flowering stage in the young flower buds which fail to develop into capsules^{10,11}. Various cultural and insecticidal management practices are followed by several workers to minimize the yield losses in linseed caused by bud fly^{5,12,13}. However, the application of insecticides increases the input cost and also leaves a residual effect in the grains. The development of resistant cultivars is one of the most effective, sustainable and widely accepted approaches to mitigate this. Most of the commercially grown linseed varieties in India are susceptible to bud fly attack^{14–16}. Therefore, potential donors for resistance need to be identified for sustainable production and to cater to rising demands for food and pharma industries¹⁷. Earlier, screening of linseed germplasm for bud fly resistance was done under deliberate late sown condition to identify potential sources of resistance^{18–20}. However, a systematic evaluation is still lacking. It is a polygenic trait²¹, hence multi-environment phenotyping is the best approach to tag the potential donors^{22–26}. Recently, we performed multi-environment screening of large collections of 2598 germplasm accessions to develop a reference set for bud fly resistance for linseed improvement and identified four genotypes as potential donors for imparting resistance²⁶. These accessions may be utilized as donors in the linseed breeding programme for developing cultivars with enhanced resistance to bud fly resistance. In the era of genomics, the use of omics resources can also be used as a tool for accelerating the molecular breeding programme²⁷. Several gene families have been characterized by previous researchers as Ranson et al.²⁸ characterized three major enzyme families such as carboxylesterases, glutathione transferases, and cytochrome P450s in the *Anopheles gambiae* genome responsible for metabolic resistance to insecticides. Yu et al.²⁹ suggested the role of Cytochrome P450 monooxygenases in hormone regulation, metabolism of xenobiotics and biosynthesis or inactivation of endogenous compounds which leads to insect resistance. The significant role of protease inhibitors (PIs) in modulating biotic stress resistance has also been reported by various workers^{30–32}. PIs play a crucial role in plant defense against predators and pathogens, with well-established functions in integrated pest management programs. While, some of the PIs are reported to regulate endogenous protein processes or act as storage proteins, many have been identified as potent defensive compounds against pests. In-depth studies have been conducted by previous workers to understand the roles and mechanisms of action of various PIs originating from three plant families: Leguminosae, Solanaceae, and Gramineae^{33–36}. These inhibitors exhibit a broad spectrum activities including suppression of pathogenic nematodes and inhibition of insect digestive enzymes. Serine, cysteine, kunitz, aspartic, and metalloproteases are the main classes of proteases targeted by these inhibitors³⁷. Transgenic crops expressing PI genes have shown resistance to a wide range of insect-pests, demonstrating the potential of these genes in integrated pest management strategies. The inhibitors within the Bowman-birk inhibitor (BBI) family, kunitz proteinase inhibitors, and cysteine proteinase inhibitors are among the well-studied PIs³⁸. They exhibit insecticidal effects against various crop pests and offer practical advantages for engineering resistance in plants^{39,40}. Additionally, the inhibitors from the potato inhibitor I and II families, such as Cathepsin-D inhibitors, are unique in their insecticidal properties and are induced upon wounding or insect damage⁴¹. Therefore, PIs may present promising opportunities for developing transgenic crops with enhanced resistance to insect pests and improved nutritional quality. Keeping the facts under consideration, the present study was planned to explore and characterize the PI candidates in the linseed genome and their possible role in modulating bud fly resistance.

Results

Identification of PI protein sequences in linseed genome

In the present investigation, 100 LuPI protein sequences were identified in the linseed (*L. usitatissimum*) genome, and downloaded from Phytozome using the Pfam IDs-pfam 00031, pfam 00280, pfam 00197, and pfam 00234 (Supplementary Table S1). These sequences were categorized based on their types and assigned the names as LuPI-01 to LuPI-100. To gain deeper insights into PI proteins, the analysis included parameters such as amino acid (aa) length, molecular weight, isoelectric point, and subcellular localization prediction for the 100 LuPI proteins. The lengths of LuPI proteins exhibited a range from 70 aa residues (LuPI-89, 90, 94) to 448 aa residues (LuPI-100), with an average length of 187 aa. Notably, a majority of the PI proteins fell within the range of 100 to 300 aa. The molecular weight spanned from 7.8 kDa (LuPI-89) to 53.92 kDa (LuPI-97), while the isoelectric point varied from 4.18 (LuPI-72) to 10.03 (LuPI-44). Sub-cellular localization predictions indicated that a significant proportion (69%) of LuPI proteins were anticipated to be in the extracellular matrix. However, some proteins were also predicted to reside in other sub-cellular compartments, including the cytoplasm (7%), nuclear

membrane shared with other organelles (8%), chloroplasts (14%), plasma membrane (10%), and mitochondria (4%), respectively.

Evolutionary tree analysis and classification of protease inhibitors

To elucidate the evolutionary relationships among 100 LuPI protein sequences, along with 15 AtPIs from *Arabidopsis* and 57 OsPIs from rice were taken for a comprehensive analysis (Fig. 1). All the LuPIs were categorized into five distinct subfamilies such as 21 Kunitz_legume inhibitors, 13 Serpins, 47 Tryp_alpha_amyl inhibitors, 7 Cystatins, and 12 Potato inhibitors-I. All subfamilies exhibited similarities with their counterparts in *Arabidopsis thaliana* and *Oryza sativa* PI subfamilies. Phylogenetic analysis revealed a bifurcated evolutionary tree, dividing into two major groups. One group comprised Tryp_alpha_amyl inhibitors and the other group comprised four PI subfamilies, revealing a closely shared origin. Subfamilies Serpin and Cystatin clustered together in one clade, while Potato inhibitor and Kunitz_legume inhibitors were placed in different clades, although they displayed a similar origin. Dissimilarity analysis provides insights into intra- and inter-subfamily variations.

Dissimilarity within sub-populations was measured as follows: Kunitz_legume inhibitor (0.63), Serpin (6.48), Tryp_alpha_amyl (2.06), Cystatin (5.42), and Potato inhibitor (1.03). Distances between subfamily groups indicated values of 4.64 (Kunitz_legume inhibitor vs. Potato inhibitor), 14.52 (Serpins vs. Tryp_alpha_amyl), 4.97 (Tryp_alpha_amyl vs. Cystatin), and 24.62 (Cystatin vs. potato inhibitor). Further group comparisons revealed

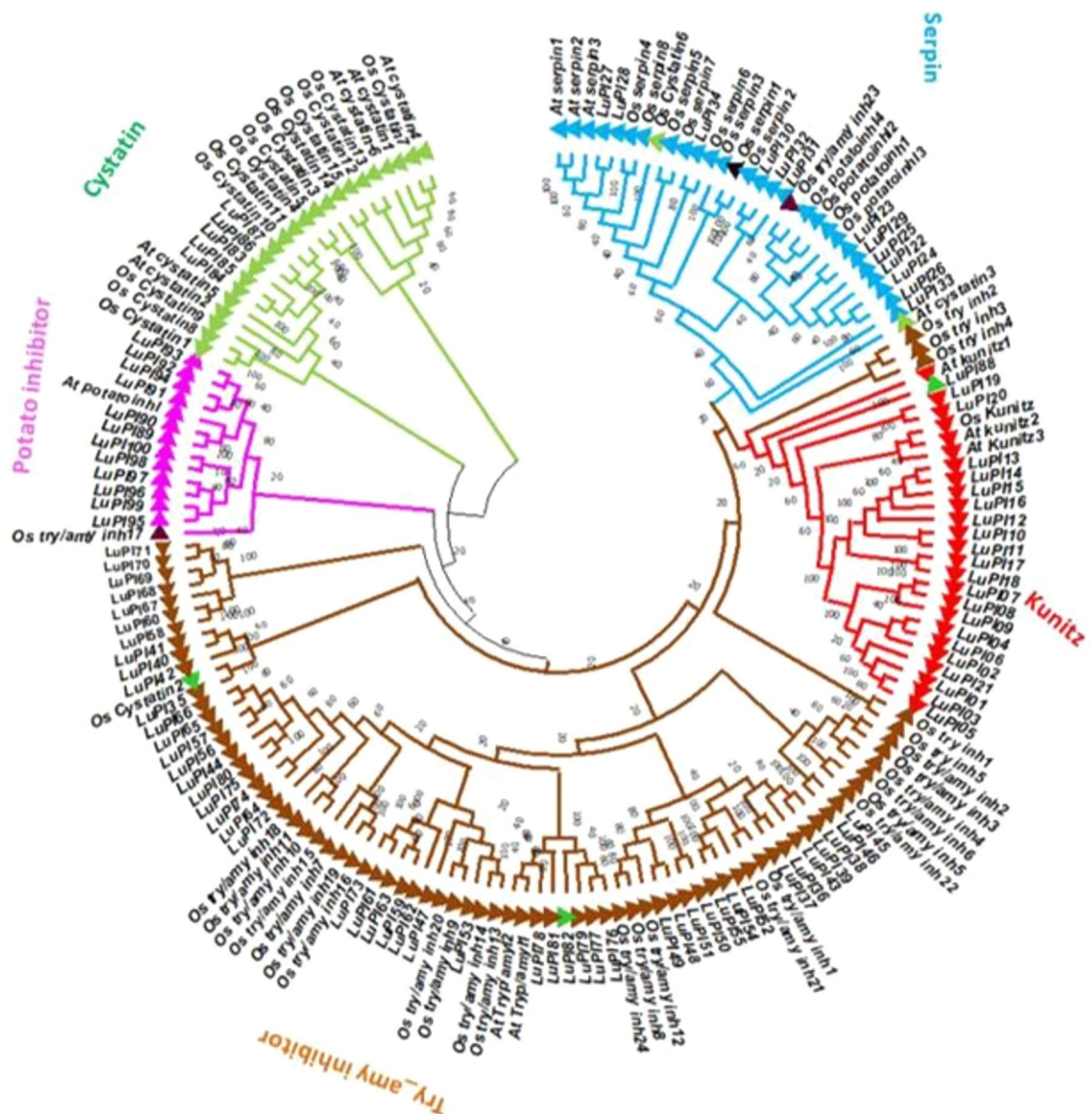


Figure 1. Evolutionary tree of LuPI genes of *Linum usitatissimum* (LuPI), *Oryza sativa* (OsPI) and *Arabidopsis thaliana* (AtPI) were used to construct the tree using MEGA 11.0 with the ML method. Subgroups Kunitz_legume inhibitors, Serpins, Tryp_alpha_amyl inhibitors, cystatins, and Potato inhibitors are indicated by red, blue, brown, green, and pink colors, respectively.

distances of 12.94 (Serpin vs. potato inhibitor), 15.47 (Tryp_alpha_amyl vs. Cystatin), and 7.43 (Tryp_alpha_amyl vs. potato inhibitor). Notably, distances between Tryp_alpha_amyl inhibitors and Potato inhibitors-I, as well as between Cystatin and Potato inhibitors-I, were determined to be 18.11 and 7.62, respectively. The relationship between Cystatin and Potato inhibitors-I exhibited a distance of 27.81. The overall diversity across the population was calculated as 8.87, with intra-population diversity as 5.75 and intra-subpopulation diversity as 3.37 (coefficient of differentiation = 0.65). Additionally, the 47 Tryp_alpha_amyl inhibitors were further classified into five types (Type 1–5). Types 1 and 2 demonstrated a mixture of Cystatins from both linseed and rice, while Type 5 Tryp_alpha_amyl members exhibited similarities with the Kunitz_legume inhibitor group.

Analysis of gene structure, conserved domain, and motifs

GSDS was utilized to illustrate the gene structure diagram of various LuPI, as depicted in (Fig. S1). Among the Kunitz_legume type of protease inhibitors, all PIs exhibited an approximately similar in size, except for *LuPI-15* and *LuPI-21*, which have two and four exons, respectively. Among the 13 Serpin-type of PIs, only four have single exon, while rest of the PIs had more than two exons. In linseed, the 46 Tryp_alpha_amyl type of PIs consisted of 14 single-exon genes, 20 with two exons, and the remaining genes possessed multiple exons. In Cystatins, all members had single exon, except for one with two exons. Similarly, all the 12-Potato inhibitor like PIs were split into two or multiple exons. The exons of all five PI subfamily; genes exhibited strikingly similar lengths and distribution, suggesting a common ancestral origin for this subfamily. Most of the LuPI candidates had genomic DNA lengths between 3 and 5 kb with some exceptions. Notably, Kunitz_legume type PIs and CPIs share similar gene structures with a single exon, while Tryp_alpha_amyl and potato inhibitor type PIs consisted of 2–6 exons. Furthermore, it was observed that most LuPI genes within the same subfamily exhibited similar exon-intron distribution patterns.

Conserved domain and motif analyses indicated that the majority of LuPIs share a similar type of arrangement of conserved domains and motifs, as illustrated in (Figs. S2, S3). Across various subfamilies of PIs, specific domains are consistently conserved in all protein sequences. Notably, Kunitz_legume-type of PIs feature two distinct domains: Kunitz_legume and STS superfamily, both abundant in glycine (G). In the Serpin type, common domains include “Serpin” and “Serpin superfamily”. Within the 46 Tryp_alpha_amyl PIs, two highly conserved domains are the AAI_LTSS superfamily and Tryp_alpha_amyl. Additionally, *LuPI-35*, *LuPI-39*, and *LuPI-41* possess an extra domain, LTP_2. CPI exhibits two conserved domains: Cystatin and CY domain, while potato inhibitor contains the major potato inhibitor superfamily domain. Some potato inhibitors also feature inhibitor_I78 superfamily and Mpv17_PMP22 superfamily domains rich in Cysteine. The representative gene structure, conserve domain and motif were presented in (Fig. 2).

To analyze the conserved motifs, ten motifs (1, 2, 4, 6, 7, 8, 9, 10) were examined in linseed genome (Figs. 3, 4, 5). Kunitz_legume-type of PIs were found to contain six motifs, while *LuPI-19* and *LuPI-20* exhibited only two motifs. In Serpin-type of PIs, motifs 6 and 7 were common, Tryp_alpha_amyl types carried motifs 8 and 9, and potato inhibitor types had motif 10. The Glycine (Gly)-rich Motif 1, crucial for the biological activity of PIs, is highly conserved among Kunitz_legume-type family members in linseed. Motif 6, consisted of 16 amino acids [DIBGTKALALPSACG], is highly conserved across all PIs (Supplementary Table S2).

The cis-acting elements analysis

Total four categories of *cis*-acting elements were identified, including light responsiveness, phytohormone responsiveness, biotic and abiotic stress responsiveness, and plant growth and development-related elements (Fig. 6). Therefore, it is speculated that the expression of LuPIs may be regulated by multiple factors. Many light-responsive elements are present in the promoters of PI genes, of which the number of G-box elements is the largest. LuPI contains a *cis*-acting element involved in phytohormone responsiveness classification. Further, analysis of the phytohormone responsiveness elements revealed that the number of elements related to MeJA was the largest, followed by abscisic acid. Around 66% of LuPI promoters contain > 6 sites for ABRE and ABA responsive elements. Abscisic acid-responsive elements were found in 76% of LuPI promoters and only *LuPI-95* contained seven sites for Abscisic acid. Furthermore, auxin (TGA-element, AuxRR-core), gibberellin (P-box, TATC-box, and GARE motif), and salicylic acid (TCA-element) are also present in LuPI promoters but ethylene responsive elements (ERE) are not detected in any LuPI promoters. Besides, the promoters also contain several types of stress responsiveness elements (SRE), including anaerobic induction elements (ARE), anoxic specific inducibility (GC-motif), low temperature (LTR), MYB drought-inducibility binding site (MBS), defense and stress (TC-rich

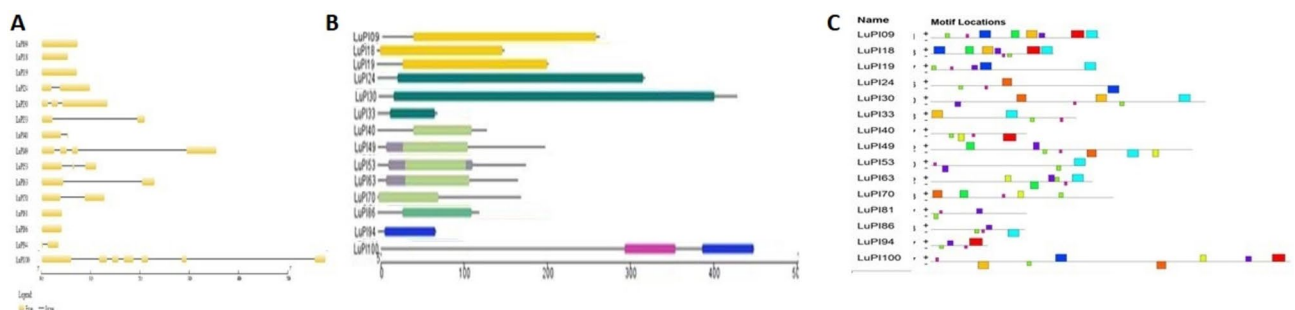


Figure 2. Intron-exon gene structure (A) domain (B) and motif (C) analyses of representative LuPI genes.

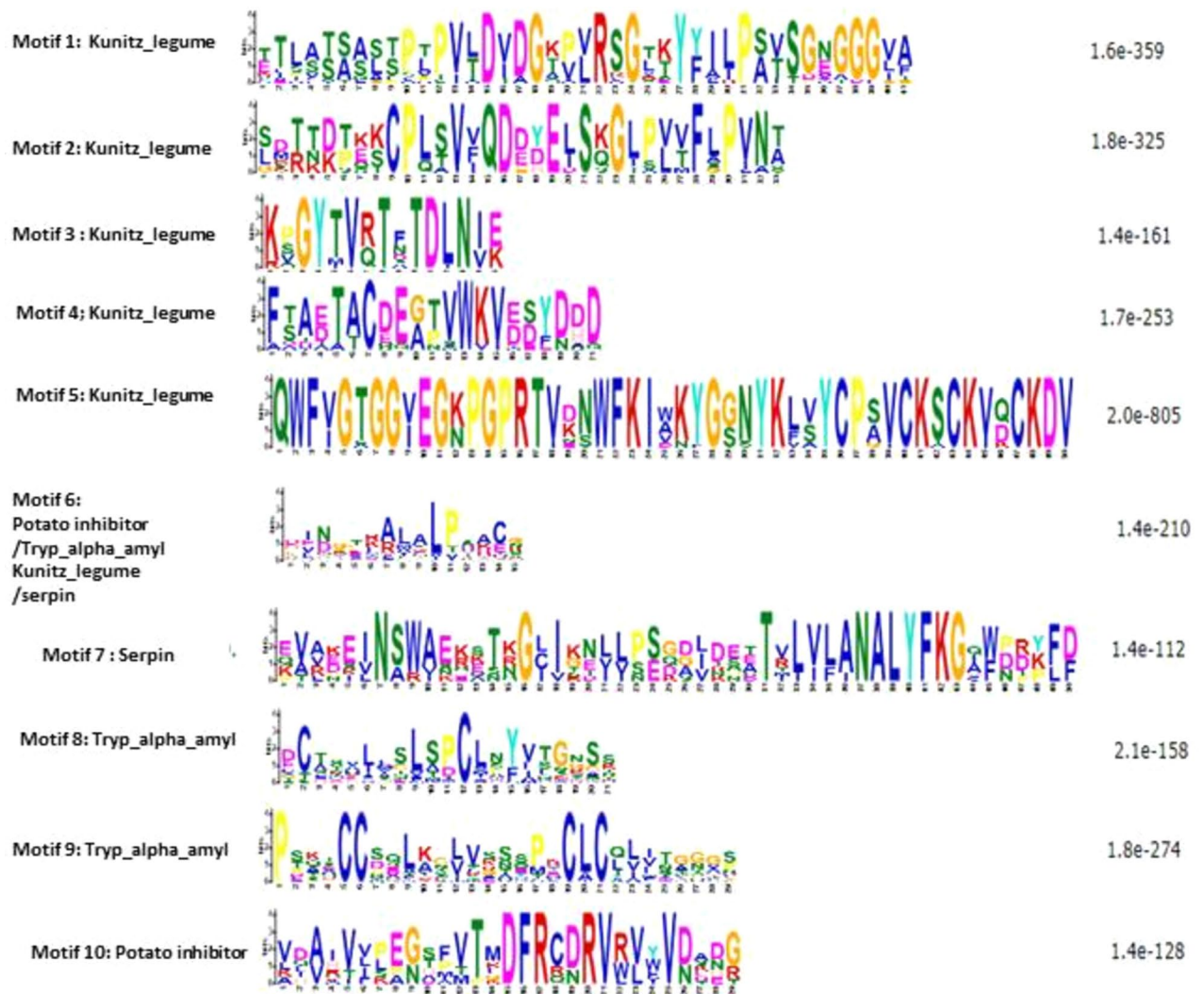


Figure 3. The conserved motifs sequences of LuPI identified by MEME.

repeats), and wound (WUN-motif) responsive elements. Additionally, plant growth and related development-related *cis*-elements in charge of meristem expression (CAT-box), circadian (circadian), endosperm expression (GCN4_motif), and Zein metabolism (O2-site) regulation were found in the promoter regions of LuPIs.

In-silico expression and qRT-PCR analysis

To ascertain the expression levels across two distinct studies, principal component analysis (PCA) was utilized. This study employed PCA to categorize gene expression during developmental stages into two clusters (Fig. S4). PCA served to diminish the data's dimensionality, facilitating the visualization of high-dimensional gene expression profiles in a more condensed space. Transcript abundance of 100 LuPIs was assessed across various developmental stages and organs, revealing notable expression during both seed and stem developmental phases (Fig. 7). The GEO data indicate that 43 LuPIs exhibited abundant transcript levels in the early stage of seed development, with 13 being highly expressed and the remainder showing moderate expression. In the second stage of seed development, 50 genes were highly expressed, including eight with particularly elevated expression. During the third stage, the number of genes expressed increased from 43 to 65, with 11 genes demonstrated the expression levels more than four folds the initiation stage. Throughout seed developmental stages, *LuPI-86*, *LuPI-87*, *LuPI-99*, and *LuPI-100* consistently showed higher expression, while *LuPI-27* and *LuPI-28* exhibited increased expression from stage 1 to stage 4. Conversely, *LuPI-50*, *LuPI-51*, *LuPI-65*, and *LuPI-75* genes maintained moderate expression across all stages, while *LuPI-52* expression decreased during developmental stages. *LuPI-54* was only expressed in the S2 stage of seed development. 34 LuPI genes were expressed at the apical shoot apex, and 40 LuPI genes were expressed at the basal shoot apex. Notably, *LuPI-84* showed high expression at the basal shoot.

To investigate the potential role of LuPI genes under bud fly infestation, we examined the expression profiles of 15 LuPI genes in a susceptible (IC0385380) and highly resistant (EC0099001) genotype, along with positive resistant controls (Neela) under control and bud fly infestation in flower buds (Fig. 8). The phenotypic data of these genotypes were presented in Table S3. For the Kunitz_legume type protease inhibitor, three genes were selected for expression profiling. The results revealed that *LuPI-19* exhibited down-regulation in all the genotypes

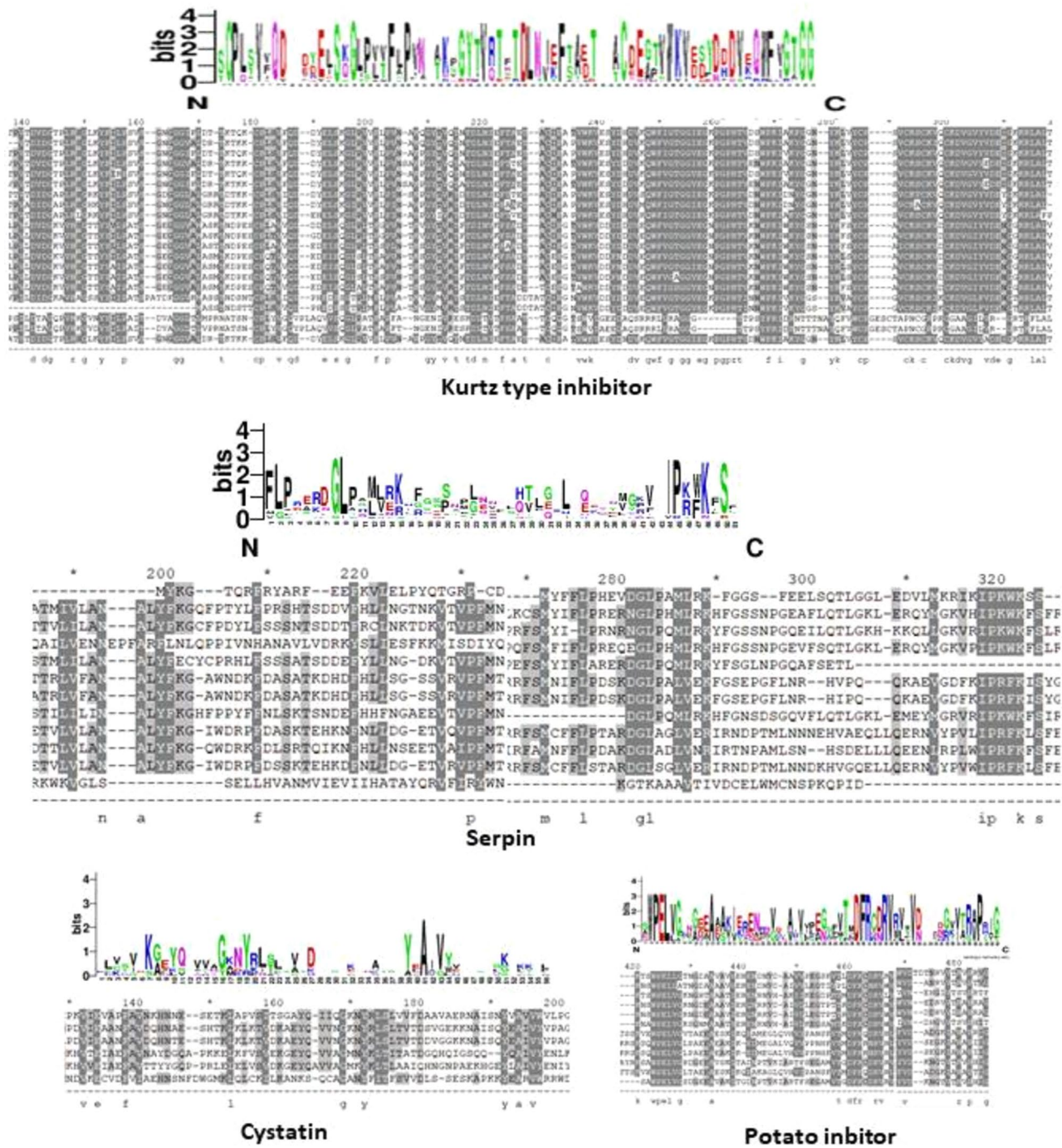


Figure 4. The architecture of conserved protein motifs of LuPIs. Different motifs are displayed shaded along with HMMER logo used for identification of Kurtz type inhibitors, serpin, cystatin, and potato inhibitors.

upon bud fly infestation. While, *LuPI-18* and *LuPI-9* showed significant up-regulation in the HR genotype (EC0099001 and Neela) upon bud fly infestation. Considering the serpin-type PI, *LuPI-24* demonstrated a significant 25-fold increased expression in EC0099001. In lineed, the predominant PI is the trypsin alpha amylase inhibitor, for which six genes i.e. *LuPI-40*, *LuPI-81*, *LuPI-49*, *LuPI-53*, *LuPI-63*, and *LuPI-70* were considered for expression profiling. All six genes showed significant expression, with *LuPI-49* exhibiting the most robust response in EC0099001 upon bud fly infestation. *LuPI-63* and *LuPI-53* also displayed significant up-regulation in HR genotype and positive control *cv. Neela*. For cystatin, only one gene was tagged and for the potato inhibitor, two genes were taken for expression profiling. Cystatin did not show a significant response, while potato inhibitor *LuPI-94* exhibited basal expression in HR upon bud fly infestation.

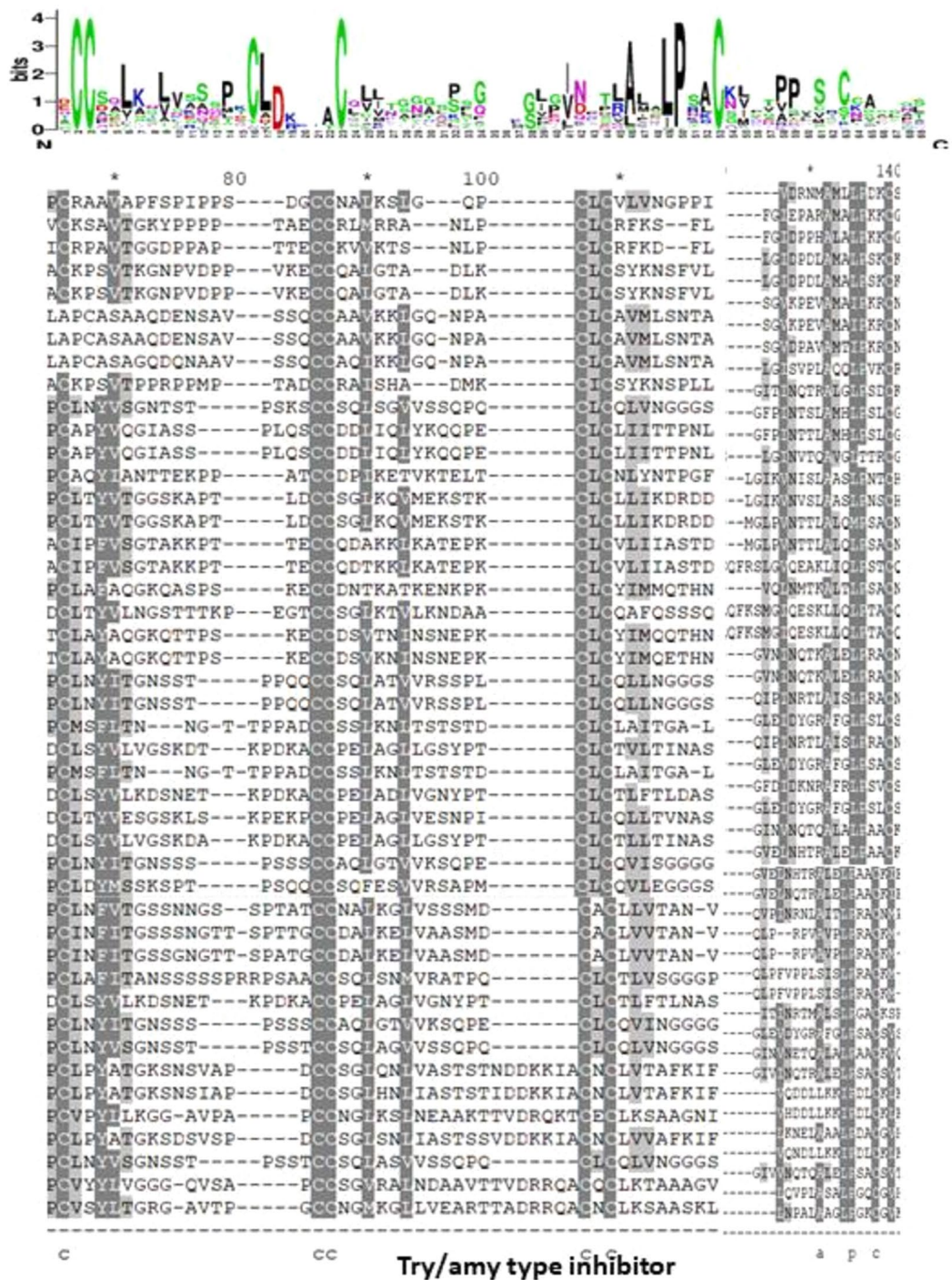


Figure 5. The architecture of conserved protein motifs of LuPI displayed shaded along with HMMER logo used for Try/amyI inhibitor identification.

Protein-protein interaction network analysis

The protein-protein interaction network of *LuPI* genes through string revealed minimal exploration in the realm of protein-protein interactions (Fig. 9). When compared with the PI genes in *Arabidopsis*, it became evident that the Tryp_alpha_ amyI inhibitor exhibited interactions exclusively with the candidates having same domain, displaying no association with other inhibitors. Distinctly, *LuPI-70* and *LuPI-40* demonstrated a comprehensive array of interactions, playing pivotal roles in diverse biological processes such as signal transduction and regulation of gene expression. Conversely, *LuPI-49*, *LuPI-53*, *LuPI-63*, and *LuPI-81* exhibited a more selective pattern of interactions. For instance, *LuPI-81* displayed co-expression with both *LuPI-49* and *LuPI-40*, while *LuPI-63* demonstrated interaction with *LuPI-49* and *LuPI-70*. This nuanced exploration sheds light on the intricate web of LuPI gene interactions, unraveling their diverse roles in biological processes.

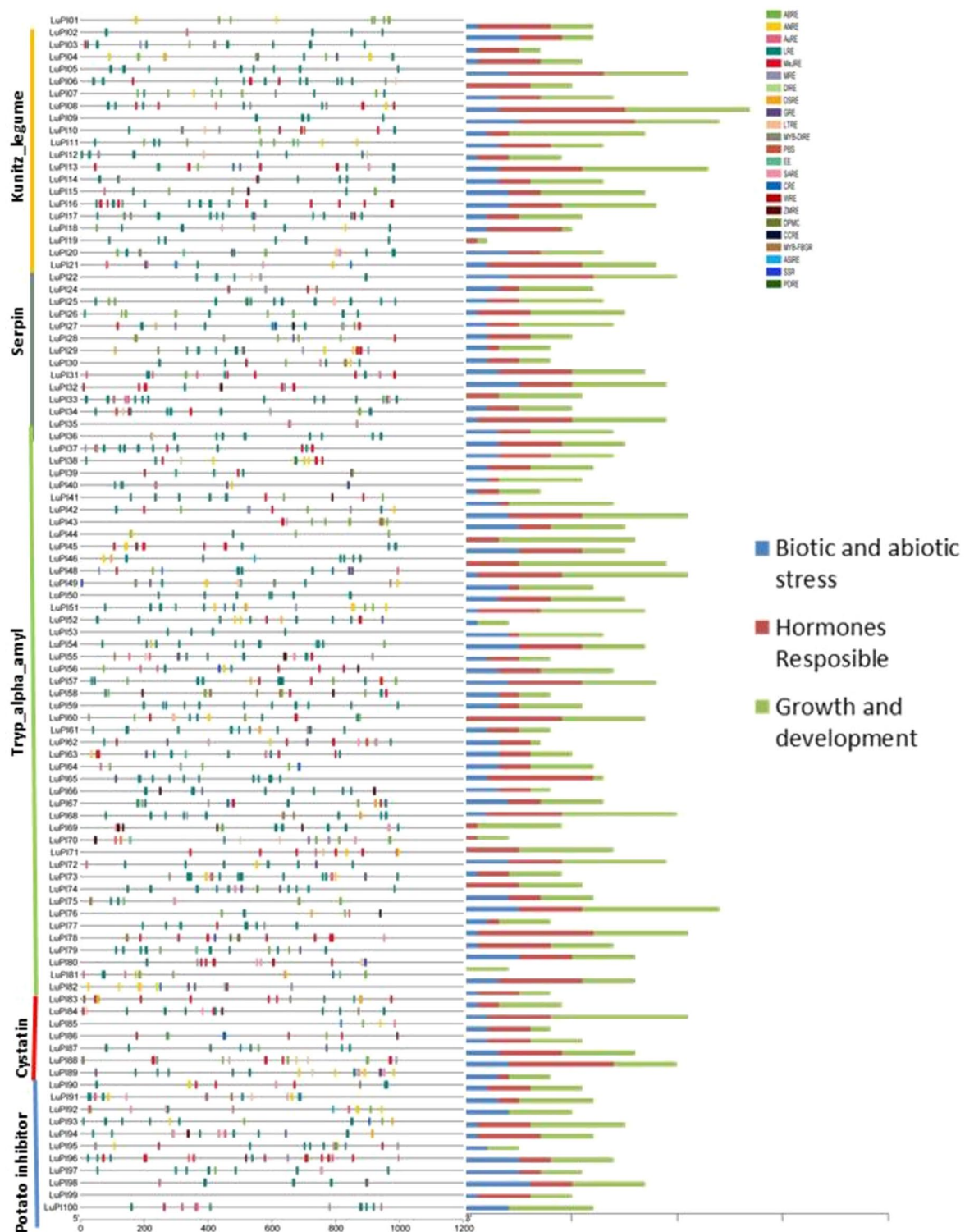


Figure 6. Distribution of *cis*-acting elements in the promoter region of LuPI genes and their functional classification.

Discussion

The PI gene family is well known for their significant role in activating defense mechanisms against several insect-pests⁴². In this investigation, 100 LuPI genes were identified in the *L. usitatissimum* genome. Subcellular localization predictions indicated that a significant proportion (69%) of LuPI proteins were anticipated to be in the extracellular matrix. However, the rest were located in the cytoplasm (7%), nuclear membrane shared with other organelles (8%), chloroplasts (14%), plasma membrane (10%), and mitochondria (4%), respectively. Fan et al.³⁴ characterized the PI gene family in tomato and predicted that 92% candidates were located as extracellular.

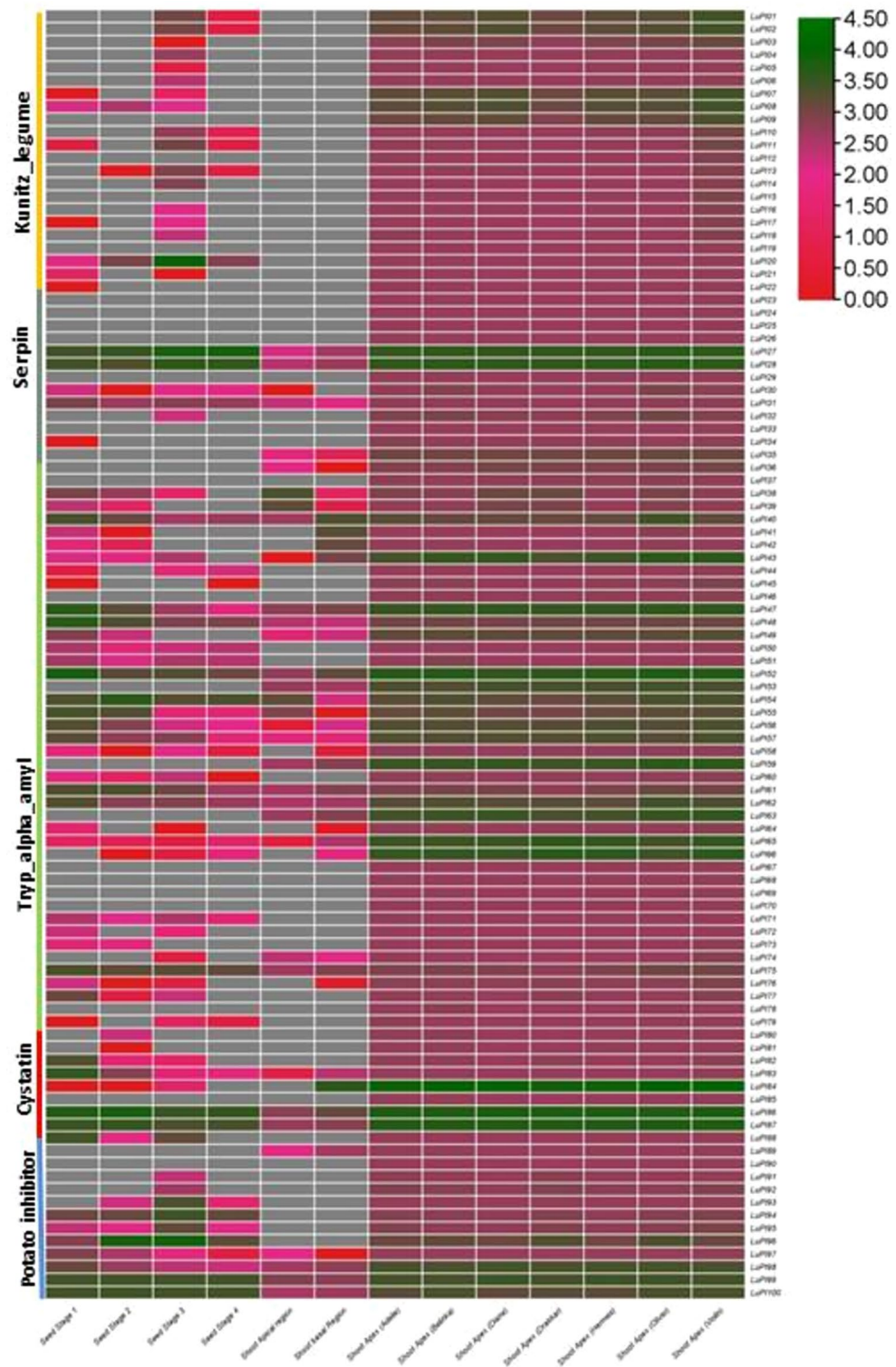


Figure 7. Heat map representation of LuPI genes across different tissues and developmental stages. The Illumina RNA-seq data were reanalyzed, and the FPKM values were log₂ transformed and heat map was generated using TB tool software. The bar at the side represents log₂ transformed values.

Rehman⁴³ characterized the Serpin gene family and noticed that most of the Serpin genes were located in chloroplast with some exceptions as they are found in the cytoplasm, endoplasmic reticulum, mitochondria, nucleus and plasma membrane. Plant PIs are a class of peptides or proteins that are known for their proteolysis

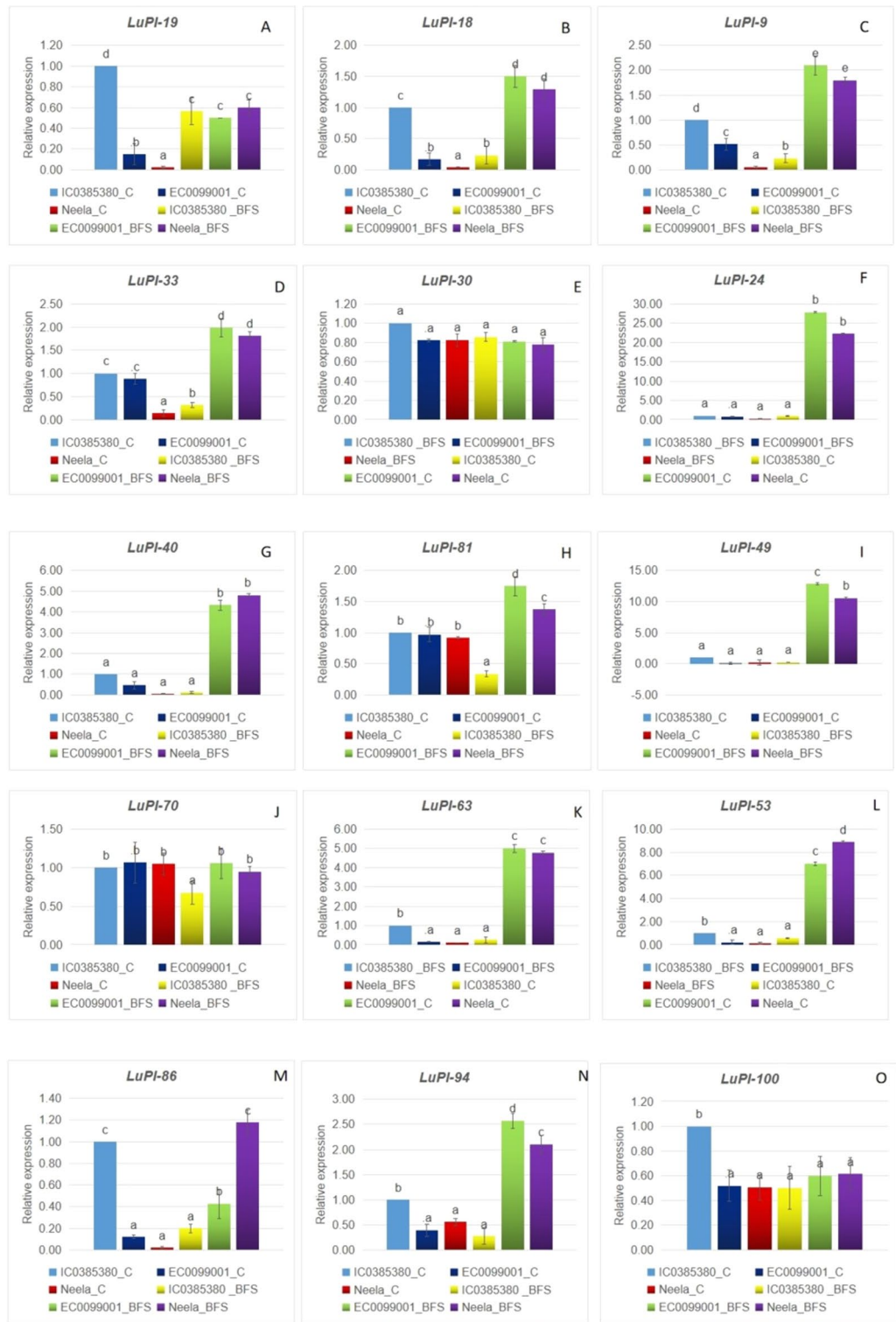


Figure 8. Relative gene expression analysis of selected LuPI genes (A) *LuPI-19*, (B) *LuPI-18*, (C) *LuPI-9*, (D) *LuPI-33*, (E) *LuPI-30*, (F) *LuPI-24*, (G) *LuPI-40*, (H) *LuPI-81*, (I) *LuPI-49*, (J) *LuPI-70*, (K) *LuPI-63*, (L) *LuPI-53*, (M) *LuPI-86*, (N) *LuPI-93*, (P) *LuPI-100* under controlled and upon bud fly infestation. The expression values were presented in fold change. The different letters showed significant differences among them.

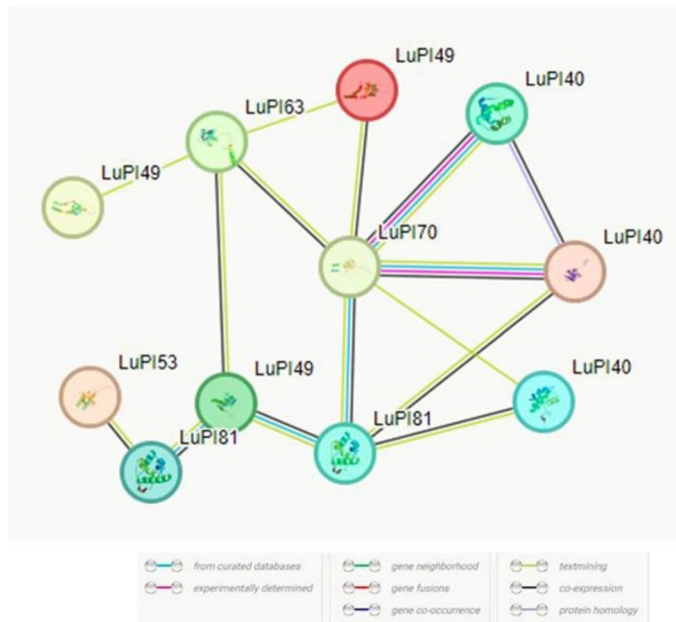


Figure 9. The network prediction of *LuPI-49* and its protein interaction based on STRING database.

activities^{33,44}. In the present investigation, 100 LuPI proteins contain at least one inhibitor domain, which are known as simple inhibitors. Therefore, they can be easily divided into different families, Kunitz_legume family, Serpin, Tryp_alpha_amyl inhibitor, Cysteine proteinase inhibitor, and Potato inhibitor according to the domain and sequence similarity. These PI proteins confer diverse functions, suggested its multiple roles in plants. The Kunitz gene family is a complex family with various PIs and inhibits serine, cysteine, and other hydrolases. The Kunitz PIs reversibly interact with their target proteases to form stable complexes and inhibit their catalytic activity in a competitive or noncompetitive manner⁴⁵. The phylogenetic tree represents the data well, as evidenced by high bootstrap branch support. This gene family is well conserved, indicating some level of functional maintenance of the gene family by natural selection. The significance of the phylogenetic relationships within the LuPI gene family offered valuable insights into the evolutionary dynamics and diversification patterns of these crucial proteins. The categorization of PI into distinct subfamilies, such as Kunitz_legume, Serpin, Tryp_alpha_amyl, Cystatin, and Potato inhibitors-I, provide a systematic framework for understanding the functional diversity in linseed. The observed similarities with AtPI and OsPIs underscore the conservation of these gene families across different plant species. The bifurcated evolutionary tree and the close origin of the Tryp_alpha_amyl group with the other four subfamilies reveal intriguing patterns of shared ancestry. The dissimilarity analysis further refines our understanding of intra and inter-subfamily variations, highlighting the diverse evolutionary trajectories among different PI groups. Notably, the identification of distinct types within the Tryp_alpha_amyl inhibitors, their relationships with Cystatins, and admixture provide an insight into our understanding of cross-species interactions and potential functional overlaps. The diversity metrics provide quantitative measures of the overall variations, offering a perspective on the genetic landscape of LuPIs. Overall, this work may serve as a foundation for future investigations into the functional roles, adaptive strategies, and evolutionary significance of LuPIs. All the LuPI genes were present in the scaffold and their chromosome location has not been assigned yet, while the PI genes in tomatoes were unevenly distributed across chromosomes³⁴. Most of the LuPI genes (48%) contained single exons and 50% had 2–4 exons. Domains and motifs are pivotal components in protein families, playing essential roles in structure, function, and evolutionary relationships. Domains with distinct structural units with specific functions contribute to the overall functionality of a protein, and their conservation across species reflects evolutionary importance. Both domains and motifs are very important in protein classification, annotation, and identification of their crucial function.

Conserved domain and motif analyses indicated that the majority of LuPIs shared similar types and signatures. The Kunitz family plays an important role against lepidopteran and coleopteran pests in various plants, including *Arabidopsis*, poplar, rice, potato, and tobacco³⁸. The Kunitz inhibitors isolated from *A. thaliana* (*AtKTI4* and *AtKTI5*) transiently expressed in *Nicotiana* plants showed their bifunctional features to inhibit cysteine- and serine-proteases present in the midgut interfering in the correct hydrolysis of dietary proteins⁴². Serpin-type inhibitors and superfamily are also found involved in pest insect-pest resistance in *Arabidopsis*^{46,47}. Clemente et al.⁴⁸ described that the Serpin worked as protease enzymes that regulate the proteolytic activity. Plant serpins are also found to enhance plant innate immunity as negative regulators of stress-induced cell death under biotic and abiotic stresses⁴³. Among the 46 Tryp_alpha_amyl types, two highly conserved domains are the AAL_LTSS superfamily and Tryp_alpha_amyl. Additionally, *LuPI-35*, *LuPI-39*, and *LuPI-41* possess an extra domain, LTP_2 had role in insect tolerance. Hamza et al.³⁵ explained that the barley protease/ α -amylase inhibitors showed a high inhibition of trypsin-like activity and it has been successfully used to improve resistance toward

Tuta absoluta. Zhao et al.⁴⁹ described that trypsin inhibitor gene products affect the growth, development and reproduction of spotted alfalfa aphids by reducing cellular enzyme activity. The mustard trypsin inhibitor affects the fertility of *Spodoptera littoralis* larvae fed on transgenic tobacco lines⁵⁰. Martinez et al.⁵¹ reviewed PhyCys and its diverse function in different physiological processes of the plant involved in the regulation of endogenous or heterologous proteases. Their defensive function is to consider PhyCys as proteins of particular value with big potential to be integrated as a new tool in insect management. The expression of barley cysteine proteinase inhibitor (Hv-CPI2) in tomato promote endogenous defense response and enhanced resistance against insect⁴⁸. Some potato inhibitors also feature inhibitor_I78 superfamily and Mpv17_PMP22 superfamily domains rich in cysteine. Wounding and UV exposure led to an increase in the expression of these two inhibitors in tomato and potato leaves are involved in plant defense against herbivorous⁴¹. The different PIs were obtained from different crop plants, like rice, barley, soybean cowpea, sweet potato and maize and were overexpressed in several plant species conferring resistance to several species of insect pests like *Nilaparvata lugens*, *Chilo suppressalis*, *Sitophilus oryzae*⁴⁶. Our results showed that the expression of LuPI genes is regulated by multiple factors. Further analysis of the phytohormone responsiveness elements revealed that the number of elements related to MeJA was the largest, followed by abscisic acid. Rehman et al.⁵² explained that ABA and MeJA-dependent signaling pathways are involved in the stimulation of the PI-II gene in *Nicotiana benthamiana*. Expression of the *UNUSUAL SERINE PROTEASE INHIBITOR (UPI)* gene is significantly induced by jasmonate, salicylic acid and abscisic acid, but is repressed by ethylene, indicating complex phytohormone regulation of *UPI* gene expression study in *Arabidopsis*⁵³. Besides, promoters also contain several types of stress responsiveness elements, including anaerobic induction (ARE), anoxic specific inducibility (GC-motif), low temperature (LTR), MYB drought-inducibility binding site (MBS), defense and stress (TC-rich repeats), and wound (WUN-motif) responsive elements presence of these *cis*-elements showed protease inhibitors (PIs) in response biotic as well as to abiotic stresses⁵⁴.

In-silico expression revealed that all 100 LuPIs were assessed across various developmental stages and organs, revealing notable expression during both seed and stem developmental phases. The LuPI genes were significantly induced upon bud fly infestation. In the present study, we validated the involvement of PIs in the linseed defense against bud fly infestation. In our investigation into the potential role of LuPI genes in linseed upon bud fly infestation, we scrutinized the expression profiles of LuPI genes in both highly susceptible and highly resistant genotypes. Among the fifteen LuPI genes studied, *LuPI-49* categorized as a Tryp_alpha_amyl type of PI, exhibited the highest expression. The protein-protein interaction revealed multi-functionality in all genes employed in the expression study, as evidenced by GO terms such as GO:0006952 (defense response), GO:0006950 (stress response), GO:0030414 (Peptidase Inhibitor Activity), and other enzymatic and peptidase activities. Importantly, *LuPI-9* and *LuPI-49* demonstrated a specific response to insect stimuli (GO:0009625) (Supplementary Table S4).

Materials and methods

Identification of protease inhibitor genes from *L. usitatissimum*

The protein, genomic, and CDS sequences from *L. usitatissimum* were accessed on 05 October, 2023 from the phytozome (<https://phytozome-next.jgi.doe.gov>). HMMER3.0 software was used to search against the *L. usitatissimum* protein sequences using the Hidden Markov model. PI protein sequences using Pfam 00031, Pfam 00280, Pfam 00197 and Pfam 00234 were downloaded from the Pfam database (<http://pfam.xfam.org/>) with E-value $\leq 1 \times 10^{-5}$ to obtain candidates, which were further submitted to conserved domain database (CDD, <https://www.ncbi.nlm.nih.gov/cdd>) for the confirmation of the existence of the PIs. ExPASy (<https://web.expasy.org/protparam/>) was used to analyze the basic physicochemical properties of PIs. Sub-cellular localization was predicted through WoLF PSORT software (<https://wolfsort.hgc.jp/>). The global sequence alignment program Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) in the EMBOSS tool was used to perform pairwise alignment of protein sequences to determine the similarity and identity between PIs.

Phylogenetic tree analysis and classification of protease inhibitor

LuPI protein sequences were compared multiple times and aligned by ClustalW and the evolutionary tree was constructed using MEGA 11.0 by the maximum likelihood (ML) method. The parameters for constructing the evolutionary tree used bootstrap values of > 1000, the Poisson correction, and pairwise deletion.

Prediction of gene structure, domain and motifs

The gene structure of LuPIs was drawn by GSDS (<http://gsds.cbi.pku.edu.cn/>). The conserved motifs of PI (10 maximum number of motifs) were analyzed using MEME suite (<http://meme-suite.org/tools/meme>) and visualized using TBtools software. The CodonW software (version 1.4.2, <http://codonw.sourceforge.net/>) was used to calculate the effective number of codons (ENC), codon adaptation index (CAI), relative synonymous codon usage (RSCU), and other codon preference parameters by MEGA 11.0 software. The presence of different domains was identified in PI protein using CDD (a conserved domain database)⁵⁴.

The *cis*-acting elements analysis

The promoter regions were used to identify the potential function of the protease inhibitor gene in *L. usitatissimum*. The 1200 bp upstream regions of all LuPI candidate genes were used to identify the *cis*-acting element using PlantCARE database⁵⁵. The positions of the promoter in the chromosome were also identified by using the TBtools programme. Adobe Illustrator was adopted for the visualization of the positions.

RNA-seq data analysis and in-silico expression

To investigate the expression profiles of LuPI genes, we analyzed transcriptome data obtained from Illumina RNA-Seq reads sourced from the NCBI GEO data set. The RNA-seq data with accession numbers GSE1130378,

GSE80718, and GSE222066 were used for *in-silico* expression of candidate genes under seed development, different stages of stem and also from seven flax cultivars (Diane, Hermes, Drakkar, Belinka, Adelie, Violin, and Oliver).

Plant materials and stress treatment

The plant material for the present study comprised three linseed genotypes namely IC03850380 (susceptible to bud fly), EC0099001 (highly resistant to bud fly), and IC0523801 (resistant check Neela). These germplasm accessions were received from the National Genebank, Indian Council of Agricultural Research–National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi, India. The seed multiplied through single plant progeny was used for screening whole germplasm collection at NGB (2598 accessions) to identify potential sources of resistance to linseed bud fly for use in the present study²⁶. The experiments for the present study were raised in two sets: One as control and another set was challenged with a linseed bud fly infestation under an insect-proof net house (ideal conditions) and under field conditions (hot spot). Ten random buds were selected to confirm the egg laying inside the buds. Six days after egg laying, the tiny flower buds were plucked from each genotype and frozen in liquid nitrogen. The young flower buds from each genotype under control were also harvested and frozen in liquid nitrogen.

RNA extraction and cDNA synthesis

The 100 mg of frozen flower bud samples from each genotype under control and stressed (challenged with bud fly) were homogenized using a mortar pestle using liquid nitrogen and subjected to RNA extraction. The RNA was extracted using the plant RNA extraction kit (RNeasy Mini Kit, Qiagen) as per the manufacturer's instructions. DNA contaminants were removed using DNase-I treatment. Further, 1 µg of pure RNA was reverse-transcribed by using the Revert Aid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Quantification of cDNA was done on a micro-volume spectrophotometer (QIAExpert, Qiagen) and normalized by 100 ng/ µL for gene expression profiling of selected candidates through qRT-PCR analysis.

Quantitative real-time PCR analysis

Fifteen gene-specific primers were designed for expression analysis (Supplementary Table S5). The PCR reactions comprised of 10 µL 2X SYBR green q-PCR master mix (Thermo Fisher Scientific), 1 µL of 10 pmol each of forward and reverse primers (Eurofins, India), 6 µL nuclease-free water, and 2 µL of cDNA. The standard cycling approach was adopted with 10 min. initial denaturation at 96 °C, 40 cycles of 30 s. denaturation at 96 °C, 60 s. annealing and extension at 60 °C. The *Actin* gene was used as a reference gene for internal control. qRT-PCR analysis was carried out using a rotor gene Q-6000 real-time PCR machine (Qiagen). Three biological replicates were taken and two technical replicates were used for expression analysis. The relative expression levels of the genes were calculated following the delta-delta CT method⁵⁶.

Protein-protein interaction

Interacting networks of LuPI proteins were integrated into the STRING (<https://www.string-db.org/> (accessed on 5 October, 2023) software, followed by an export of the co-expression network data from STRING through KEGG⁵⁷, which was further calculated using Microsoft Excel 2019.

Ethical approval

The plant collection and use were in accordance with all the relevant guidelines. The seeds of germplasm accessions used in the present study were deposited in National Genebank (NGB), Indian Council of Agricultural Research–National Bureau of Plant Genetic Resources (ICAR-NBPGR), New Delhi, India.

Sharing of germplasm

Linseed is not part of Annexure I crops of the International treaty on plant genetic resources for food and agriculture (ITPGFRA) and therefore not part of the multi-lateral system (MLS) of access and benefit sharing. Eligible users from within India can request and receive the material as per the terms and conditions and rates (http://www.nbpgr.ernet.in/Portals/6/services/Fee_Stru_2023.pdf) following the material transfer agreement (<http://www.nbpgr.ernet.in/Downloadfile.aspx?EntryId=7379>). Collaborative projects can be the best way to use the linseed core collection following the material transfer agreement for International Bilateral Exchange under Collaborative Research Programmes (<http://www.nbpgr.ernet.in/Downloadfile.aspx?EntryId=7380>).

Conclusions

The present investigation systematically identified and characterized the 100 LuPI genes in the linseed genome elucidating their diverse subcellular localizations. Phylogenetic analysis revealed the conservation and evolutionary dynamics of the PI gene family, offering a framework for assessing functional diversity through subfamily categorization. The gene structure unveiled varied exon–intron organizations, while analysis of *cis*-acting elements in promoter regions indicated potential regulation by factors such as phytohormones and stress responsiveness. The activation of PI systems in response to insect-pest invasions is of great importance as it allows for the monitoring of plant conditions and prompt actions to reduce pest-induced harm. Through the integration of PI systems with technologies for monitoring insect-pest activity, linseed industries can optimize their production procedures and reduce losses caused by pest infestations. *In-silico* expression, the analysis revealed significant expression of Tryp_amyl inhibitor after challenging with bud fly. Tryp_amyl inhibitor as the largest inhibitor in linseed and all six genes especially *LuPI-49*, *LuPI-63* performed well under bud fly stress highlighting the

crucial role in defense mechanisms. This study suggests that Tryp₂ amyl inhibitor are good candidates and worthy while breeding linseed for bud fly resistance. This comprehensive study establishes a foundation resource for the structural, functional, and evolutionary dimensions of LuPI genes.

Data availability

All the data generated in this experiment are presented in the manuscript and its supplementary materials. The in-silico expression was performed on GSE222066, GSE130378, and GSE80718.

Accessions

The accessions are available in the NGB as per the respective National IDs (IC03850380 (susceptible to bud fly), EC0099001 (highly resistant to bud fly), and IC0523801 (resistant check Neela)). Further inquiries can be directed to the corresponding author (Dr. Vikender Kaur).

Received: 6 January 2024; Accepted: 30 July 2024

Published online: 02 August 2024

References

- Kajla, P., Sharma, A. & Sood, D. R. Flaxseed—a potential functional food source. *J. Food Sci. Technol.* **52**, 1857–1871 (2015).
- Singh, R. B. & Singh, R. N. Date of sowing and varieties for root-rot and wilt complex in linseed (*Linum usitatissimum* L.). *Indian J. Agric. Sci.* **81**, 287–289 (2011).
- FAOSTAT (2022) Available at: <https://www.fao.org/faostat/en/#data/QCL> (August 5 2022).
- AICRP report 2. <https://aicrp.icar.gov.in/linseed/pc-massage/> (May 13 2024).
- Malik, Y. P., Hussain, K., Singh, S. V. & Srivastava, R. L. Development of management module for bud fly *Dasyneura lini* in linseed. *Indian J. Entomol.* **62**, 260–269 (2000).
- Malik, Y. P. Estimation of economic threshold level for bud fly, *Dasyneura lini* Barnes in linseed (*Linum usitatissimum* L.). *J. Oilseeds Res.* **22**, 100–102 (2005).
- Malik, Y. P. Yield losses due to bud fly (*Dasyneura lini* Barnes) in linseed. *J. Oilseeds Res.* **23**, 363 (2006).
- Paliwal, S. *et al.* Molecular advances to combat different biotic and abiotic stresses in linseed (*Linum usitatissimum* L.): A comprehensive review. *Genes* **14**(7), 1461 (2023).
- Biswas, G. C. & Das, G. P. Insect and mite pests diversity in the oilseed crops ecosystems in Bangladesh. *Bangladesh J. Zool.* **39**, 235–244 (2011).
- Singh, P. K. & Yadav, D. K. Inheritance of resistance to bud fly infestation in linseed (*Linum usitatissimum* L.). *Indian J. Genet. Plant Breed.* **75**, 386–388 (2015).
- Kumar, V. *et al.* Biophysical and morphological basis of resistance against linseed (*Linum usitatissimum* L.) bud fly (*Dasyneura lini* Barnes). *Indian J. Plant Genet. Resour.* **36**(3), 382–386 (2023).
- Singh, B., Katiyar, R. R., Malik, Y. P. & Pandey, N. D. Influence of sowing dates and fertilizer levels on the infestation of linseed bud fly (*Dasyneura lini* Barnes). *Indian J. Entomol.* **53**, 291–297 (1991).
- Malik, Y. P., Husain, K. & Alam, K. Impact of plant spacing and fertilizer application on linseed and infestation of bud fly (*Dasyneura lini* Barnes). *J. Oilseeds Res.* **25**, 106–107 (2008).
- Kanwar, P. *et al.* Seasonal incidence of linseed bud fly (*Dasyneura lini* Barnes) on different varieties in relation to weather parameters. *Pharm. Innov. J.* **11**(4), 963–965 (2022).
- Reddy, M. P., Reddy, B. R. & Maheshwari, J. J. Screening of linseed genotypes for resistance against bud fly, alternaria and powdery mildew, genetic parameters for yield components in linseed (*Linum usitatissimum* L.). *Int. J. Curr. Microbiol. Appl. Sci.* **2**, 267–276 (2013).
- Ekka, R. K., Mandal, S., Meena, R. S. & Padamshali, S. Some physical basis of host plant resistance in linseed (*Linum usitatissimum* L.) against bud fly, *Dasyneura lini* (Barnes). *J. Pharmacogn. Phytochem.* **7**, 2010–2015 (2018).
- Verma, P. S. *et al.* Linseed: An oilseed crop harnessing the power of nature. *Indian Farm.* **73**(7), 22–25 (2023).
- Pal, R. K. & Singh, R. Screening of linseed germplasms against bud fly (*Dasyneura lini* Barnes). *Int. J. Plant Prot.* **3**, 410–411 (2010).
- Pal, R. & Malik, Y. P. Histological basis of resistance in linseed against bud fly (*Dasyneura lini* Barnes) in central Uttar Pradesh. *Int. J. Agric. Invent.* **5**, 25–30 (2020).
- Pal, S., Shrivastava, J. L. & Pande, N. D. Effect of different dates of sowing on the incidence of *Dasyneura lini* Barnes. *Indian J. Entomol.* **40**, 433–435 (1978).
- Singh, P. K. & Yadav, D. K. Inheritance of resistance to bud fly infestation in linseed (*Linum usitatissimum* L.). *Indian J. Genet. Plant Breed.* **75**(3), 386–388 (2015).
- Singh, C. M. *et al.* Microsatellite-based association mapping for agronomic traits in mungbean (*Vigna radiata* L. Wilczek). *J. Genet.* **100**, 87 (2021).
- Kaur, V. *et al.* Diversity of *Linum* genetic resources in global genebanks: From agro-morphological characterisation to novel genomic technologies—A review. *Front. Nutr.* **10**, 1165580 (2023).
- Kaur, V. *et al.* Multi-environment phenotyping of linseed (*Linum usitatissimum* L.) germplasm for morphological and seed quality traits to assemble a core collection. *Ind. Crops Prod.* **206**, 117657 (2023).
- Tripathi, A. *et al.* Identification of potential sources of mungbean yellow mosaic India virus resistance in black gram (*Vigna mungo*) and expression of antioxidants and R-genes modulating resistance response in cultivated and its two wild relatives. *Plant Breed.* **142**(5), 668–681 (2023).
- Nair, B. *et al.* Multi-environment screening of *Linum* germplasm collection for dissecting the potential of bud fly (*Dasyneura lini* Barnes) resistance and assembling a reference set for efficient utilization in genetic improvement. *Ind. Crops Prod.* **207**, 117743 (2023).
- Yadav, B. *et al.* Integrated omics approaches for flax improvement under abiotic and biotic stress: Current status and future prospects. *Front. Plant Sci.* **13**, 931275 (2022).
- Ranson, H. *et al.* Evolution of supergene families associated with insecticide resistance. *Sci.* **298**, 179–181 (2002).
- Yu, L. *et al.* Characterization and expression of the cytochrome P450 gene family in diamondback moth, *Plutella xylostella* (L.). *Sci. Rep.* **5**, 8952 (2015).
- Shamsi, A. & Bano, B. Journey of cystatins from being mere thiol protease inhibitors to at heart of many pathological conditions. *Int. J. Biol. Macromol.* **102**, 674–693 (2017).
- Carrillo, L. *et al.* Expression of a barley cystatin gene in maize enhances resistance against phytophagous mites by altering their cysteine-proteases. *Plant Cell Rep.* **30**(1), 101–112 (2011).
- Moloi, S. J. & Ngara, R. The roles of plant proteases and protease inhibitors in drought response: A review. *Front. Plant Sci.* **14**, 1165845 (2023).

33. Sultana, M. S. *et al.* Proteinase inhibitors in legume herbivore defense: From natural to genetically engineered protectants. *Plant Cell Rep.* **41**(2), 293–305 (2022).
34. Fan, Y. *et al.* Genome-wide identification and expression analysis of the protease inhibitor gene families in tomato. *Genes* **11**(1), 1 (2019).
35. Hamza, R. *et al.* Expression of two barley proteinase inhibitors in tomato promotes endogenous defensive response and enhances resistance to *Tuta absoluta*. *BMC Plant Biol.* **18**, 24 (2018).
36. Purwar, S. *et al.* Basal expression studies of cystatins during specific growth stages of wheat spikes for defining their possible role in differential and stage dependent immunity against karnal bunt (*Tilletia indica*). *Mol. Biol. Rep.* **37**, 1377–1389 (2010).
37. Gatehouse, J. A. Prospects for using proteinase inhibitors to protect transgenic plants against attack by herbivorous insects. *Curr. Protein Pept. Sci.* **12**(5), 409–416 (2011).
38. Arnaiz, A. *et al.* Arabidopsis kunitz trypsin inhibitors in defense against spider mites. *Front. Plant Sci.* **9**, 390273 (2018).
39. Turra, D. & Lorito, M. Potato type I and II proteinase inhibitors: Modulating plant physiology and host resistance. *Curr. Protein Pept. Sci.* **12**, 374–385 (2011).
40. Turrà, D. *et al.* Heterologous expression of PKPI and Pin1 proteinase inhibitors enhances plant fitness and broad-spectrum resistance to biotic threats. *Front. Plant Sci.* **11**, 500975 (2020).
41. Dunse, K. M. *et al.* Coexpression of potato type I and II proteinase inhibitors gives cotton plants protection against insect damage in the field. *Proc. Natl. Acad. Sci. USA* **107**(34), 15011–15015 (2010).
42. Rustgi, S., Boex-Fontvieille, E., Reinbothe, C., von Wettstein, D. & Reinbothe, S. The complex world of plant protease inhibitors: Insights into a kunitz-type cysteine protease inhibitor of *Arabidopsis thaliana*. *Commun. Integr. Biol.* **11**, e1368599 (2018).
43. Rehman, S. *et al.* Genome wide identification and comparative analysis of the serpin gene family in brachypodium and barley. *Plants* **9**(11), 1439 (2020).
44. Hellinger, R. & Gruber, C. W. Peptide-based protease inhibitors from plants. *Drug Discov. Today* **24**(9), 1877–1889 (2019).
45. Alvarez-Alfageme, F. *et al.* Potential use of a serpin from *Arabidopsis* for pest control. *PLoS ONE* **6**(5), e20278 (2011).
46. Irving, J. A. *et al.* Phylogeny of the serpin superfamily: Implications of patterns of amino acid conservation for structure and function. *Genome Res.* **10**, 1845–1864 (2004).
47. Ferreira, M. M. *et al.* Plant serpins: Potential inhibitors of serine and cysteine proteases with multiple functions. *Plants* **12**, 3619 (2023).
48. Clemente, M. *et al.* Plant serine protease inhibitors: Biotechnology application in agriculture and molecular farming. *Int. J. Mol. Sci.* **20**(6), 1345 (2019).
49. Zhao, H. *et al.* Inhibitory effects of plant trypsin inhibitors *Msti-94* and *Msti-16* on *Therioaphis trifolii* (Monell) (Homoptera: Aphididae) in Alfalfa. *Insects* **10**(6), 154 (2019).
50. De Leo, F. & Gallerani, R. The mustard trypsin inhibitor 2 affects the fertility of *Spodoptera littoralis* larvae fed on transgenic plants. *Insect Biochem. Mol. Biol.* **32**(5), 489–496 (2002).
51. Martinez, M. Phytocystatins: Defense proteins against phytophagous insects and acari. *Int. J. Mol. Sci.* **17**(10), 1747 (2016).
52. Rehman, S. *et al.* Expression analysis of proteinase inhibitor-II under OsRGLP2 promoter in response to wounding and signaling molecules in transgenic *Nicotiana benthamiana*. *3 Biotech.* **8**(1), 51 (2018).
53. Laluk, K. & Mengiste, T. The *Arabidopsis* extracellular unusual serine protease inhibitor functions in resistance to necrotrophic fungi and insect herbivory. *Plant J.* **68**(3), 480–494 (2011).
54. Marchler-Bauer, A. *et al.* CDD: A conserved domain database for protein classification. *Nucleic Acids Res.* **33**, D192–D196 (2005).
55. Lescot, M. *et al.* PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Res.* **30**(1), 325–327 (2002).
56. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2(-\Delta\Delta C(T))$ method. *Methods* **25**(4), 402–408 (2001).
57. Kanehisa, M. Towards understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**(11), 1947–1951. <https://doi.org/10.1002/pro.3715> (2019).

Author contributions

CMS, BKS, VK, and BN conceived the idea and planned the work; SP, R, AP, and SS retrieved the data from the database; CMS and SP analyzed the data; SP drafted the manuscript, BN, BKS, VK edited the manuscript. All authors read the manuscript and approved it for publication.

Funding

Department of Biotechnology, Ministry of Science and Technology, India, No. BT/Ag/Network/Linseed/2019–20.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-68943-1>.

Correspondence and requests for materials should be addressed to B.K.S. or V.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024