Bulletin of Biological and Allied Sciences Research
***ISSN:* 2521-0092**
www.bbasr.org
DOI: https://doi.org/10.64013/bbasr.v2026i1.117
*Bull. Biol. All. Sci. Res., Volume, **11**: 117*
Original Research Article ***Open Access***

GENOME-WIDE ANALYSIS OF PATHOGENESIS-RELATED PROTEIN 1 (PR1) GENES REVEALS STRUCTURAL AND EXPRESSION DIVERSITY IN *PISUM SATIVUM*

SHERAZI SHUH[1], SHAFIQ M[2], AHMED H[1], AHMAD J[3], HAMMAD M[2], SANAM A[1], REHAN M[1], TUFAIL MT[1], HAYYAT Q[1]

[1]*Department of Plant Breeding and Genetics, Faculty of Agricultural Sciences, University of the Punjab, P.O BOX. 54590, Lahore, Pakistan*
[2]*Department of Horticulture, Faculty of Agriculture Sciences, University of the Punjab, Lahore, Pakistan*
[3]*Department of Agronomy, Faculty of Agricultural Sciences, University of the Punjab, Lahore, Pakistan*
**Correspondence Author Email Address: syedhaseebulhassansherazi@gmail.com*

(Received, 7th May 2025, Accepted 29th January 2026, Published 17th February 2026)

Abstract *Pathogenesis-related gene PR1 acts as an essential element in the plant defensive mechanism and is made during the process of systemic acquired resistance (SAR) and salicylic-induced signaling. The genome-wide analysis of the PR1 gene in Pisum sativum has been conducted to analyze the functionality, composition, and evolutionary status of this gene. Based on the fact that CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1) domain, fifteen PsPR1 genes were identified. Complete CAP domain has been found on these proteins, which ensures that all belong to the PR1 family. There is a mild difference in the size of protein, the molecular mass, as well as the point of isoelectric activity showed by the physicochemical study. On the other hand, subcellular localization predictions revealed that there is predominant extracellular targeting, with the secretory nature of PR-1 proteins, in multiple cellular compartments. The Fabaceae family has a lineage-specific increase of PR-1 genes, according to phylogenetic research with several dicot species. This suggests functional conservation with known Arabidopsis thaliana PR-1 homologs involved in SAR. The majority of PsPR1 genes are either intron-less or have a single intron, supporting fast transcriptional activation during stress, according to conserved motif and gene structure analyses that revealed strong evolutionary conservation. Strong purifying selection was shown by Ka/Ks analysis of duplicated gene pairs, highlighting PR-1 genes' physiological importance in pea defence. Numerous cis-regulatory regions linked to hormone signalling, stress response, and transcription factor binding were found by promoter analysis, suggesting intricate regulatory control. PsPR1 genes were found to be distributed unevenly by chromosomal mapping, with tandem duplication being the main cause of the substantial clustering on chromosome 1. Tandem and segmental duplication events both contributed to the formation of the gene family, according to synteny analysis. PsPR1.1, PsPR1.3, and PsPR1.9 were found to be significantly regulated genes by expression profiling during heat stress, especially in the heat-tolerant cultivar Akra Chaitra. Overall, this study identifies intriguing possibilities for enhancing heat stress resilience and offers a solid foundation for comprehending the functional evolution and stress-responsive activities of PR-1 genes in Pisum sativum.*

[Citation: Sherazi, S.H.U.H., Shafiq, M., Ahmed, H., Ahmad, J., Hammad, M., Sanam, A., Rehan, M., Tufail, M.T., Hayyat, Q. (2026). Genome-wide analysis of pathogenesis-related protein 1 (pr1) genes reveals structural and expression diversity in *Pisum sativum. Bull. Biol. All. Sci. Res.* **11***: 117. doi:* https://doi.org/10.64013/bbasr.v2026i1.117]

Keywords: *PR-1 gene family*; *Pisum sativum*; *systemic acquired resistance; genome-wide analysis; gene duplication and evolution; heat stress response*

Introduction

Pea (*Pisum sativum*) is a member of the Leguminosae family, classified under the subfamily Faboideae and tribe Fabeae. Owing to its ability to fix atmospheric nitrogen, pea provides significant ecological benefits by supporting low-input agricultural systems and functioning effectively as a break crop, thereby reducing dependence on external inputs (Smýkal et al., 2012). Peas, particularly yellow and green cotyledon types commonly referred to as dry, smooth, or field peas, are the naturally dried seeds of *Pisum sativum* L. and are cultivated worldwide for both human consumption and animal feed (Dahl et al., 2012). Peas are cultivated worldwide and are considered an important component of the human diet. Globally, Canada ranks as the leading producer of peas, followed by China, Russia, and India (Wu et al., 2023). Plant immunity comprises a complex network of defense mechanisms that function to counteract pathogenic microorganisms, thereby limiting infection and protecting plant tissues from damage (Fan et al., 2025). Plant immunity relies on two main defense layers. The first, known as pattern recognition–based immunity (PRBI), provides an early and broad response against invading pathogens, while the second, effector-triggered immunity (ETI), activates a stronger and more targeted defense

mechanism (Fan et al., 2025). In addition, pathogenesis-related (PR) genes encode PR proteins that work together to strengthen plant resistance against pathogen invasion (Fan et al., 2025). PR proteins are well-characterized and belong to multiple structurally and functionally distinct families. Their levels change significantly in response to pathogen infection, wounding, or treatments with elicitors and chemical inducers such as salicylic acid, ethylene, and jasmonic acid (Anuradha et al., 2022). CAP proteins form an evolutionarily conserved superfamily that is widely expressed and found across bacteria, fungi, plants, and animals (Han et al., 2023). CAP proteins are involved in a diverse range of biological processes, including immune defense in both mammals and plants, pathogen virulence, sperm maturation and fertilization, venom toxicity, as well as prostate and brain cancers (Schneiter and Di Pietro, 2013). Most CAP proteins are secreted glycoproteins that remain highly stable in the extracellular environment across a broad range of conditions (Schneiter and Di Pietro, 2013).

Materials and Methods

Comprehensive Identification of PR-1 Genes in *Pisum sativum*

Pathogenesis-Related 1 (PR-1) genes were identified throughout the *Pisum sativum* genome. The UniProt database (https://www.uniprot.org/). provided PR-1 protein sequences with the conserved CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1) domain. BLAST searches were performed using these sequences against the P. sativum reference genome found in the Phytozome database (https://phytozome-next.jgi.doe.gov/). Non-redundant candidate PR-1 proteins were kept for additional analysis after useless sequences were eliminated.

Verification of CAP Domain

The National Centre for Biotechnology Information (NCBI) Conserved Domain Database (CDD) search tool (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). was used to evaluate all PR-1 protein sequences. Only proteins with a full CAP domain were regarded as genuine members of the PR-1 family.

Gene Expression Data

The gene expression datasets were obtained from the Gene Expression Omnibus (GEO) database of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/geo/). Publicly available expression profiles relevant to the genes of interest were selected based on experimental design and data completeness. Raw or processed expression data were downloaded, and genes corresponding to the study objectives were extracted for further analysis. Statistical analysis of gene expression data was performed to evaluate differences among experimental groups. Analysis of variance (ANOVA) was conducted using Statistix software version 8.1. The retrieved data were arranged and evaluated for suitability for parametric testing before analysis. Differences among means were considered statistically significant according to standard probability thresholds.

Phylogenetic Analysis

Phylogenetics is the field that explores the evolutionary relationships among organisms (Brinkman and Leipe, 2001). Phylogenetic analysis provides a way to infer or estimate these evolutionary relationships (Brinkman and Leipe, 2001). The Phytozome database (https://phytozome-next.jgi.doe.gov/). was used to obtain protein sequences from several crop species. MEGA software was used to create a phylogenetic tree from these sequences, and the tree was exported in Newick format. iTOL (https://itol.embl.de/) was then used to show and annotate the evolutionary relationships for an understandable and instructive graphical depiction.

Physicochemical Properties of PR-1 Proteins

The ExPASy ProtParam tool (https://web.expasy.org/protparam/) was used to determine the amino acid length, molecular weight (MW), and theoretical isoelectric point (pI) of the identified P. sativum PR-1 proteins.

Prediction of Subcellular Localization

Recent rapid advances in genomic and proteomic research have led to the accumulation of vast amounts of DNA and protein sequence data in public databases (Yu et al., 2006). This makes it increasingly important for computational biologists to develop effective tools that can efficiently extract relevant biological information from sequences for functional annotation (Yu et al., 2006). Since a protein's function is closely linked to its subcellular localization, accurately predicting this localization can greatly aid in characterizing expressed sequences with unknown functions (Yu et al., 2006). The WoLF PSORT system (https://wolfpsort.hgc.jp/), which forecasts protein localization based on amino acid composition, signal peptides, and recognized sorting motifs, was used to predict the subcellular localization of PR-1 proteins.

Gene Structure Analysis

Genome-wide analyses of protein structural patterns have proven useful for uncovering both functional and evolutionary relationships (Liu et al., 2002). The exon–intron architecture of PR-1 genes was analyzed using the Gene Structure Display Server (GSDS v2.0) (https://gsds.gao-lab.org/). Submission of Genomic DNA sequences and related coding sequences was done to visualize gene structures and to interpret structural variations among PR-1 gene family members.

Conserved Motif and Protein Domain Analysis

The MEME Suite (https://meme-suite.org/meme/tools/meme) was used to find conserved motifs among PR-1 proteins. The analysis was done using default parameters, according to the standard number of motifs normally used in gene family analysis. Identified motifs were subsequently analyzed to determine their distribution and

conservation across PR-1 proteins. Conserved domain analysis of the protein sequences was carried out using the NCBI Conserved Domain Database (CDD) via the CD-Search program (https://ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) with default parameters. Identified domains were then visualized using TBtools, which was used to generate schematic diagrams showing the domain organization.

Cis-Regulatory Element Analysis

The PlantCARE database (https://bioinformatics.psb.ugent.be/webtools/plantcare/html/) was used to examine 2 kb upstream promoter sequences from each PR-1 gene's translation start site to look into possible regulatory mechanisms controlling PR-1 gene production. Based on their roles in hormone signalling, developmental regulation, and stress response, cis-regulatory elements were categorized.

Ka/Ks Analysis and Estimation of Divergence Time

To assess evolutionary constraints acting on duplicated PR-1 gene pairs, nonsynonymous (Ka) and synonymous (Ks) substitution rates were calculated. $T = Ks / 2\lambda$ was used to calculate the divergence time (T), where λ represents the neutral substitution rate for *Pisum sativum*, obtained from previously published studies. The selection pressure on PR-1 genes was predicted using the Ka/Ks ratio.

Chromosomal Localization of PR-1 Genes

For the positions and distribution, the MG2C v2.1 tool (http://mg2c.iask.in/mg2c_v2.1/) was used to visualize the patterns of PR-1 genes on chromosomes that were discovered using genome annotation data.

Synteny

The Phytozome database (https://phytozome-next.jgi.doe.gov/) provided the genomic information needed for synteny analysis. Genome sequences and gene annotation files were downloaded and used to identify syntenic relationships based on conserved gene order and collinearity among the selected genomes. Synteny blocks were visualized using TBtools with default settings, and figures were optimized for publication quality.

Discussions

Gene Expression Dynamics of PsPR1

PR proteins are produced in plants in response to pathogen attack or other stress conditions. They are grouped into 17 distinct families, labeled PR1 through PR17 [3]. Among them, PR1 proteins are known for their antimicrobial activity, PR5, PR12, PR13, and PR14 possess antifungal properties, and PR10 can degrade viral RNA (Wufuerjiang et al., 2025). A differential expression study throughout the entire genome was carried out to assess the transcriptional responsiveness of PsPR1 gene family members under treatment conditions. Statistical assessment using F-values and corresponding p-values revealed that only a subset of PsPR1 genes exhibited significant treatment-dependent expression changes. Among the analyzed genes, PsPR1.1, PsPR1.3, and PsPR1.9 were identified as differentially expressed genes (DEGs). PsPR1.1 showed a significant response with an F-value of 7.97 ($p = 0.018$), while PsPR1.3 exhibited a comparable level of induction ($F = 9.24$, $p = 0.013$). Notably, PsPR1.9 demonstrated the strongest transcriptional response among all family members, with a markedly high F-value (17.10) and a highly significant p-value ($p = 0.002$), indicating robust sensitivity to the applied treatment. In contrast, the majority of PsPR1 genes, including PsPR1.2, PsPR1.4–PsPR1.8, PsPR1.10, PsPR1.12, and PsPR1.13, did not show statistically significant expression changes ($p > 0.05$) and were therefore classified as non-DEGs. Despite the lack of statistical significance, several genes, such as PsPR1.12, PsPR1.13, and PsPR1.14, were transcriptionally detectable, exhibiting low to moderate basal expression levels. PsPR1.14 displayed a relatively higher F-value (4.05) compared to other non-DEGs; however, the response did not reach statistical significance ($p = 0.100$). Overall, these findings suggest functional divergence within the PsPR1 gene family, where only specific members—particularly PsPR1.9—are strongly responsive to treatment conditions. This selective induction implies that distinct PsPR1 genes may play specialized roles in treatment-mediated defense signaling, while others may contribute to basal or constitutive defense mechanisms.

PsPR1 Gene	F value (Treatment)	P value	Significance	DEG Status	Remarks
PsPR1.1	7.97	0.018	*	DEG	—
PsPR1.2	1.57	0.266	ns	Non-DEG	—
PsPR1.3	9.24	0.013	*	DEG	—
PsPR1.4	1.01	0.362	ns	Non-DEG	—
PsPR1.5	0.66	0.453	ns	Non-DEG	—
PsPR1.6	0.42	0.546	ns	Non-DEG	—
PsPR1.7	0.37	0.567	ns	Non-DEG	—
PsPR1.8	1.00	0.363	ns	Non-DEG	—
PsPR1.9	17.10	0.002	**	DEG	Strongest response
PsPR1.10	0.83	0.404	ns	Non-DEG	—
PsPR1.12	0.37	0.567	ns	Non-DEG	Detected, low expression
PsPR1.13	1.00	0.363	ns	Non-DEG	Detected, minimal expression
PsPR1.14	4.05	0.100	ns	Non-DEG	Detected, moderate expression

Table 1: ANOVA results for PR1 genes in *Pisum sativum* at a significance level of p = 0.05

Phylogenetic Analysis of PR1 Gene Family

Some plants exhibit resistance to the majority of pathogens, a phenomenon referred to as non-host resistance (Ali et al., 2025). Plants defend against pathogen attack by activating a multilayered immune system capable of detecting and neutralizing a wide range of invaders (Ali et al., 2025). For example, plants defend themselves against pathogens using a two-layered immune system, which includes pattern-triggered immunity (PTI) as the first line of defense and effector-triggered immunity (ETI) as a stronger, more targeted response (Ali et al., 2025). PR1 protein sequences from *Pisum sativum* were used to create a thorough phylogenetic tree and some dicot species, including Arabidopsis thaliana, Lens culinaris, Capsicum annum, Medicago truncatula, Helianthus annuus, and Cucumis sativus. The *Pisum sativum* PR1 genes were distributed across the tree rather than being confined to a single lineage. A prominent clade had the PR1 sequences from LcPR1, MtPR1, and PsPR1 showing a lineage specific expansion within Fabaceae e.g. PsPR1.7, PsPR1.12, LcPR1.13, LcPR1.10, and MtPR1.10, MtPR1.14, MtPR1.16. A major clade contained PR-1 sequences from Arabidopsis (AtPR1.1, AtPR1.3, AtPR1.5, AtPR1.11, AtPR1.13), Sunflower (HaPR1.1, HaPR1.3, HaPR1.9), Medicago (MtPR1.19, MtPR1.21), and Cucumber (CsPR1.11, CsPR1.12, CsPR1.14, CsPR1.15). These homologs are previously associated with systemic acquired resistance (SAR) and extracellular antimicrobial secretion, implying that pea members located within these clades may perform equivalent roles.

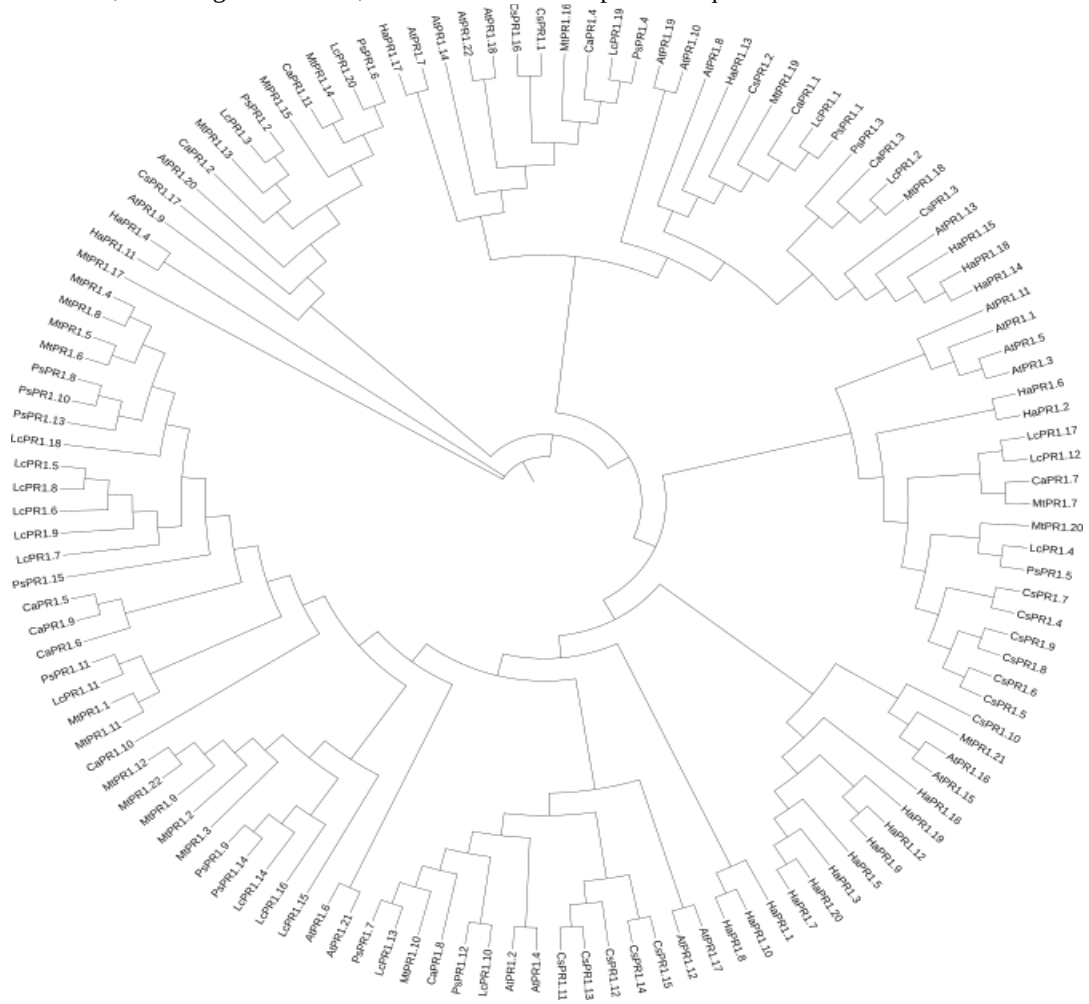

Figure 1: Phylogenetic tree of PR1 genes in *Pisum sativum* in comparison with *Arabidopsis thaliana, Cicer arietinum, Cucumis sativus, Lens culinaris, Helianthus annuus*, and *Medicago truncatula*

Gene	Chromosome	Start Position	End Position	Orientation	MW (Da)	pI	CDS Length (bp)	E value	Align Length
PsPR1.1	chr4	24,223,287	24,224,752	Forward	24,045.04	6.41	912 bp	1.36E-132	206
PsPR1.2	chr4	198,282,562	198,283,081	Forward	19,160.80	8.16	831 bp	7.42E-131	173
PsPR1.3	chr4	24,228,717	24,229,350	Reverse	24,105.30	8.39	927 bp	3.48E-162	211

PsPR1.4	chr6	98,915,049	98,917,134	Reverse	36,274.50	8.07	1,401 bp	0	317
PsPR1.5	chr7	178,879,022	178,879,502	Forward	17,988.97	4.31	639 bp	7.36E-119	160
PsPR1.6	chr2	455,833,495	455,834,044	Reverse	19,941.84	6.88	762 bp	3.33E-128	183
PsPR1.7	chr1	388,708,734	388,709,376	Reverse	23,805.79	8.18	918 bp	5.42E-161	214
PsPR1.8	chr1	451,885,414	451,885,912	Reverse	18,003.23	5.63	642 bp	4.35E-109	166
PsPR1.9	chr1	452,356,529	452,357,051	Reverse	18,727.08	8.47	678 bp	2.66E-129	174
PsPR1.10	chr1	451,867,927	451,869,870	Reverse	23,342.92	5.40	936 bp	7.16E-139	217
PsPR1.11	chr1	452,180,835	452,181,327	Reverse	18,039.44	7.57	651 bp	2.59E-123	164
PsPR1.12	chr1	451,775,905	451,776,379	Reverse	17,506.20	5.05	624 bp	1.59E-118	158
PsPR1.13	chr1	451,885,414	451,885,912	Reverse	18,069.37	5.44	642 bp	1.33E-101	166
PsPR1.14	chr1	452,272,014	452,272,536	Reverse	18,825.23	8.84	675 bp	1.73E-129	174
PsPR1.15	chr1	452,236,664	452,237,216	Reverse	20,183.83	8.14	708 bp	6.51E-139	184

Table 2: Physicochemical properties of PR1 genes in *Pisum sativum*

Physiochemical Properties

Subcellular Localization Predicted for PsPR1 Proteins

PR1 proteins are the most abundantly produced members of the PR protein family in response to pathogen attack (Pečenková et al., 2017). PR1 family members are highly conserved across plant species, and homologs have also been identified in fungi, insects, and vertebrates, including humans (Pečenková et al., 2017). Together, these proteins constitute a superfamily of secreted proteins known as CAP, named after cysteine-rich secretory proteins (CRISP), antigen 5, and PR1 proteins (Pečenková et al., 2017). Although PR1 proteins have been studied extensively, their precise function remains unclear (Pečenková et al., 2017). They are thought to possess antifungal activity and may also contribute to host defense by participating in salicylic acid (SA)-mediated signaling and hypersensitive response (HR)-related cell death (Pečenková et al., 2017). This analysis showed that PsPR1 proteins get distributed across multiple cellular compartments, but the most frequent localization site was extracellular compartments specially for the proteins named PsPR1.5, PsPR1.11, and PsPR1.9. This was in accordance with the canonical secretory nature of PR1 proteins. It indicates that evolution has diversified PR1 targeting multiple cellular compartments rather than a single secretion route. In addition, several proteins showed predicted transfer to the chloroplast, nucleus, endoplasmic reticulum, plasma membrane, cytosol, and vacuole. PsPr1.6 and PsPr1.15 notably showed vacuolar localization, contributing to storage, stress buffering or storage immunity, and plastid-related targeting, indicating more food production. Cytoplasm-nucleus shared distribution shows that some defense signaling pathways are involved in transcriptional regulations.

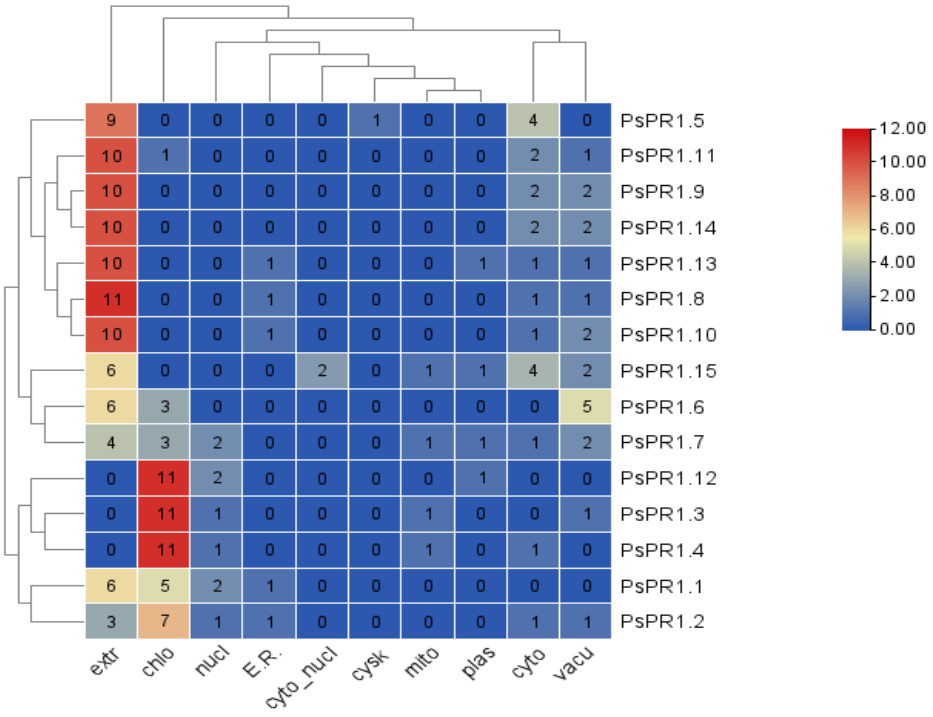

Figure 2: Subcellular localization of PR1 genes in *Pisum sativum*

Gene Structure Organization

The activation of innate immunity and systemic acquired resistance (SAR) is accompanied by the expression and accumulation of pathogenesis-related (PR) proteins (Wang et al., 2022). Upon pathogen attack, PR proteins accumulate to high levels, helping the plant combat invading pathogens (Wang et al., 2022). Gene structure analysis of PsPR1 in *Pisum sativum* compared with Arabidopsis thaliana, Capsicum annum, and Medicago truncatula revealed that most PsPR1 genes were intronless or contained only a single intron. The coding sequences were also uninterrupted. This simplicity of exon organization tells that the genes show rapid response during stress which is also consistent with the PR1 genes of other plant species. Some things to be noted here are that PsPR1.1 had a long downstream untranslated region (UTR), while PSPR1.4 had a long coding region. These variations indicate structiural complexity and neofunctionalization as discussed before.

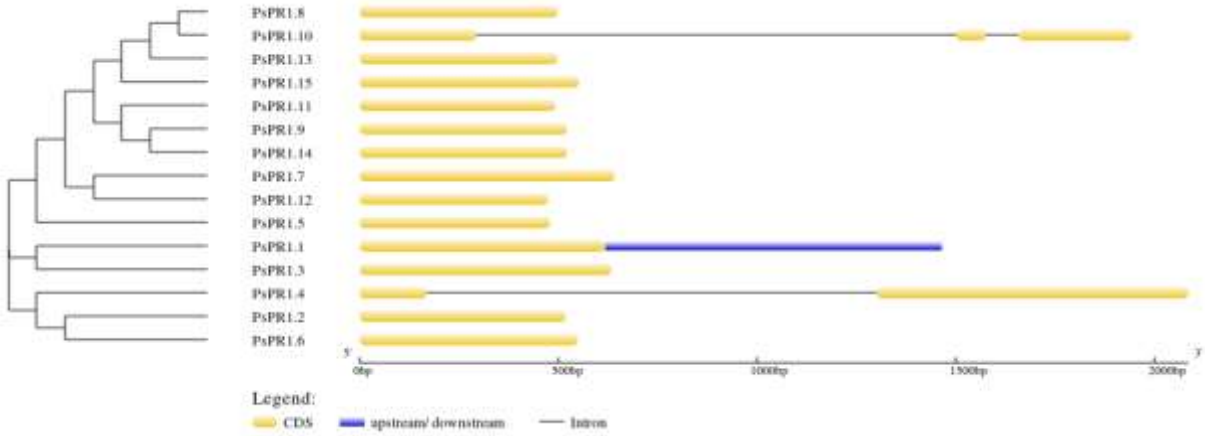

Figure 3: Gene structure analysis showing the intron–exon organization of PR1 genes

Conserved Motifs and Protein domianins

Selective protein-protein interactions play a crucial role in cellular functions and are often mediated by protein domains that recognize specific sequences within target proteins, known as Short Linear Motifs (SLiMs) (Ren et al., 2008). A significant advance in SLiM identification came with a peptide library-based approach, which allows mapping of the sequence motifs recognized by SH2 domains without requiring prior knowledge of their in vivo interaction sites (Ren et al., 2008). Comparable peptide library experiments have also been conducted to identify the motifs recognized by other protein domains (Ren et al., 2008). Motifs identified through polypeptide library screening have demonstrated a strong correspondence with previously reported domain interaction sites (Ren et al., 2008). These analyses were done to explore structural conservation in PsPR1 proteins, and 10 such motifs were identified. Motifs 7,5,4,6,2 and 1 were shared by most members, showing strong conservation through evolution. All proteins had the CAP domain. PsPR1.4 additionally showed extended CAP superfamily features, indicating structural specialization, possibly neofunctionalization.

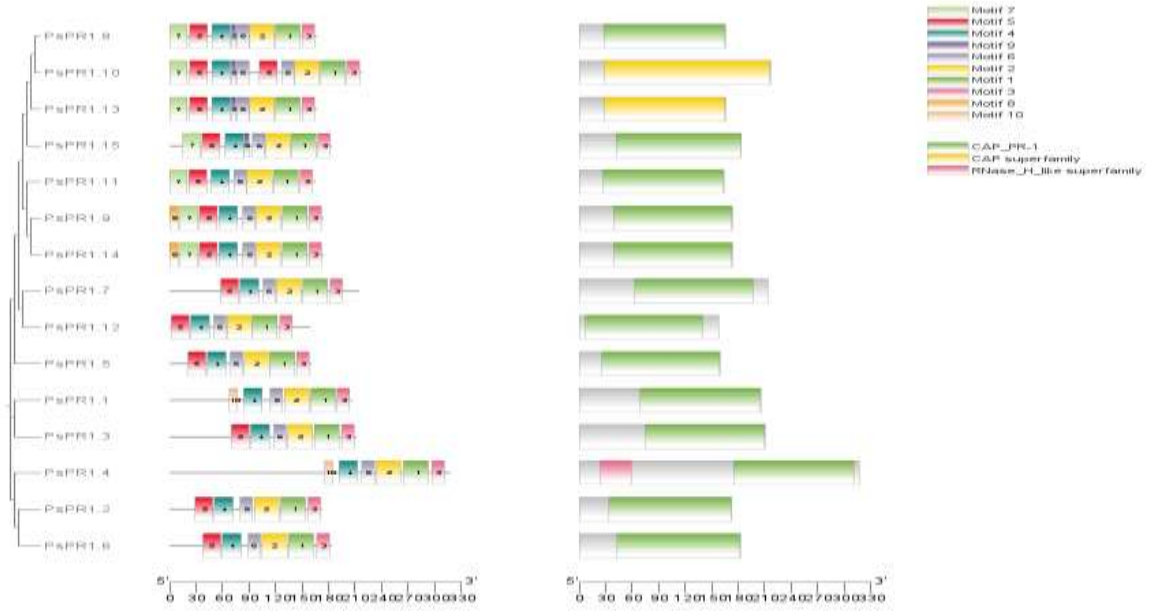

Figure 4: Analysis of conserved motifs and functional domains in *PsPR1* genes

Cis-Regulatory Element Profiling

Cis-regulatory sequences determine the timing, location, and level of gene expression (Wufuerjiang et al., 2025). Central to this regulation is the recruitment of transcription factors (TFs), which bind to regulatory sequences such as promoters, enhancers, repressors, and insulators (Wufuerjiang et al., 2025). These target sequences are distributed throughout the DNA. Likely influenced by the secondary structure of looped DNA within the nucleus, cis-regulatory elements have been observed as far as approximately 10^6 base pairs away from a gene's transcription start site (Wufuerjiang et al., 2025). Spatiotemporal and cell-type–specific gene expression patterns are regulated by DNA sequences called cis-regulatory elements (CREs) [G] (Preissl et al., 2023). The 2kb upstream promoter regions of all PsPR1 genes were analyzed, which revealed regulatory elements associated with biotic, abiotic, hormonal, and developmental responses. All PsPR1 promoters had conserved motifs, including TATA box and CAAT box in high abundance, showing a high level of transcription activation and regulation. Many hormone-responsive cis-elements were detected. It shows that PR1 genes are active in signaling pathways. These included ABRE (ABA-responsive element) associated with abscisic acid signaling, ERE (ethylene-responsive element) associated with stress and pathogen defense, Gibberellin responsiveness is linked to GARE-motif and TATC-box, auxin regulation is linked to AuxRR-core and TGA-element, and methyl jasmonate responsiveness is linked to CGTCA-motif and TGACG-motif. Many stress-responsive cis-elements were also detected, including MBS (MYB binding site) associated with drought inducibility, W-box, a binding site for WRKY transcription factors known to regulate defense genes, ARE (anaerobic responsive element), STRE (stress-responsive element), TC-rich repeats, associated with defense and stress responsiveness. Light responsive cis-elements were G-box, Box-4, GT1-motif, TCT-motif, and I-box showing light dependent regulation of defense genes. Binding factors such as MYB, MYC, WRKY and bZIP were found in high frequency.

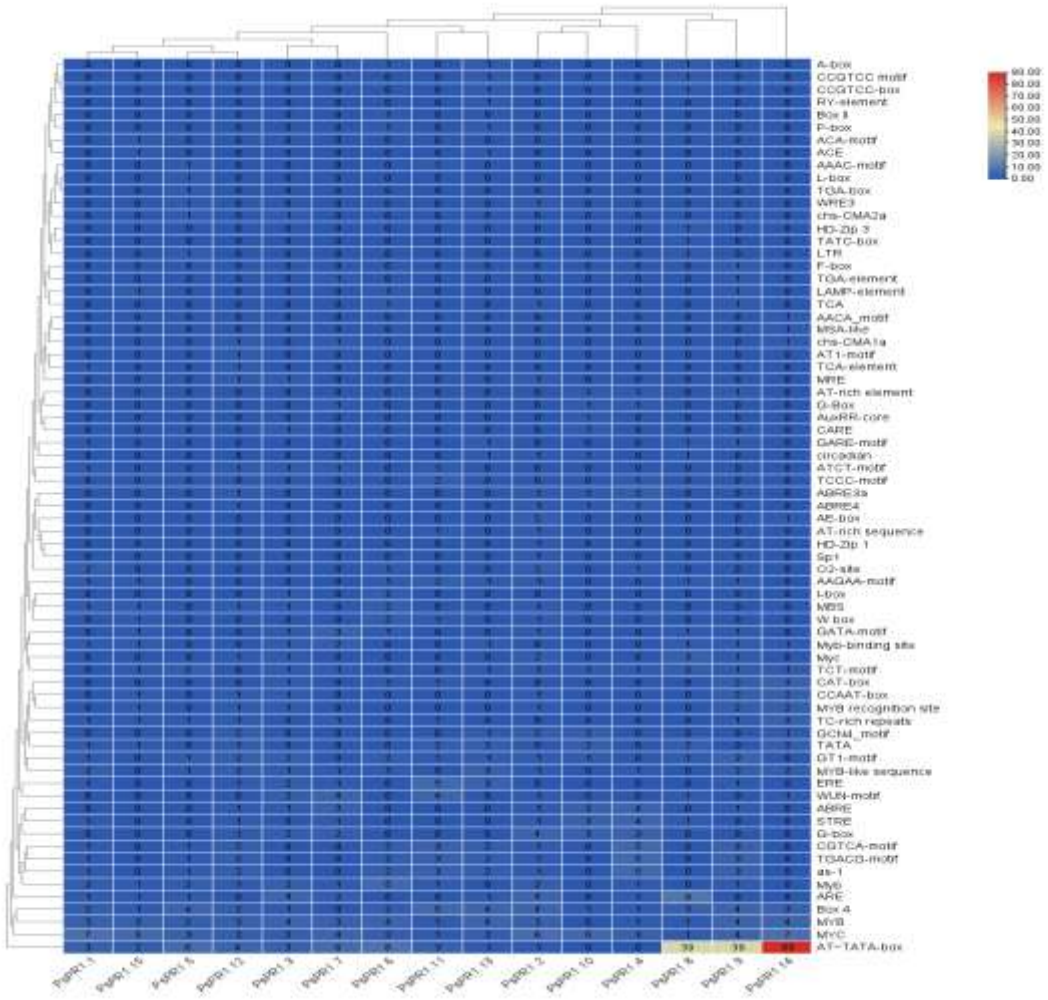

Figure 5: Heatmap illustrating the cis-regulatory elements of PsPR1 genes

Gene Duplication and Evolutionary Selection Pressure

Ka/Ks analysis was performed, and three duplicated gene pairs were identified that are: PsPR1.8–PsPR1.10, PsPR1.10–PsPR1.13, and PsPR1.9–PsPR1.14. All of them showed Ka/Ks < 1, which means that the family has gone under very strong purifying selection during their evolutionary history. Since the role of PR1 proteins is conserved, it is likely essential for plant defense in *Pisum sativum*.

Seq_1	Seq_2	Ka	Ks	Ka/Ks
PsPR1.8	PsPR1.10	0.018334	0.04756	0.385486
PsPR1.9	PsPR1.14	0.005044	0.051229	0.098463
PsPR1.10	PsPR1.13	0.021008	0.047409	0.44312

Table 3: Analysis of synonymous (Ks) and non-synonymous (Ka) substitution rates of PsPR1 genes

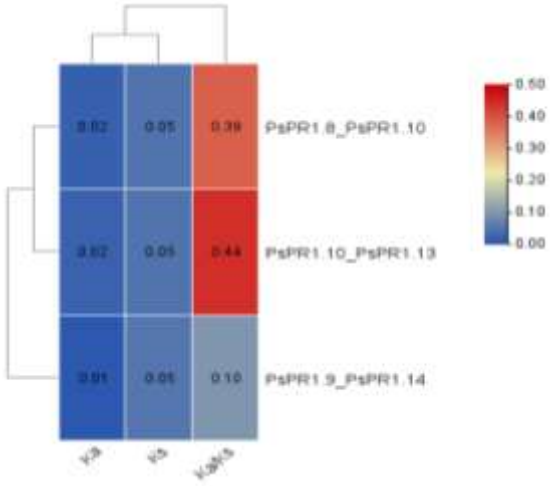

Figure 6: Heatmap depicting the synonymous (Ks) and non-synonymous (Ka) substitution ratios of PsPR1 genes

Chromosomal Localization of PsPR1 genes

The fundamental concept of marker-assisted selection is to utilize marker–trait associations to guide the development of new cultivars and breeding lines, particularly during the selection of parental pairs for crossing and the identification of desirable hybrids (Pavlenko and Pikunova, 2024). PsPR1 genes were found to be irregularly distributed across the *Pisum sativum* genome by chromosomal mapping. 15 of the genes were distributed on only 5 chromosomes, namely chromosome 1, 2, 4, 6, and 7. Chromosome 1 had 9 PsPR1 genes from PsPR1.7 to PsPR1.15 means that PsPR1 genes in *Pisum sativum* are clustered mainly on chromosome 1. This clustering shows duplication-driven expansion in the gene family. This clustering is also seen in Arabidopsis thaliana, showing an evolutionary relationship. In contrast, chromosomes 2, 6, and 7 had only 1 PsPR1 gene on each of them. Overall, chromosomal mapping analysis demonstrates that the PsPR1 gene family exhibits a highly uneven distribution, with chromosome 1 acting as the principal reservoir for PR-1 genes in *Pisum sativum*.

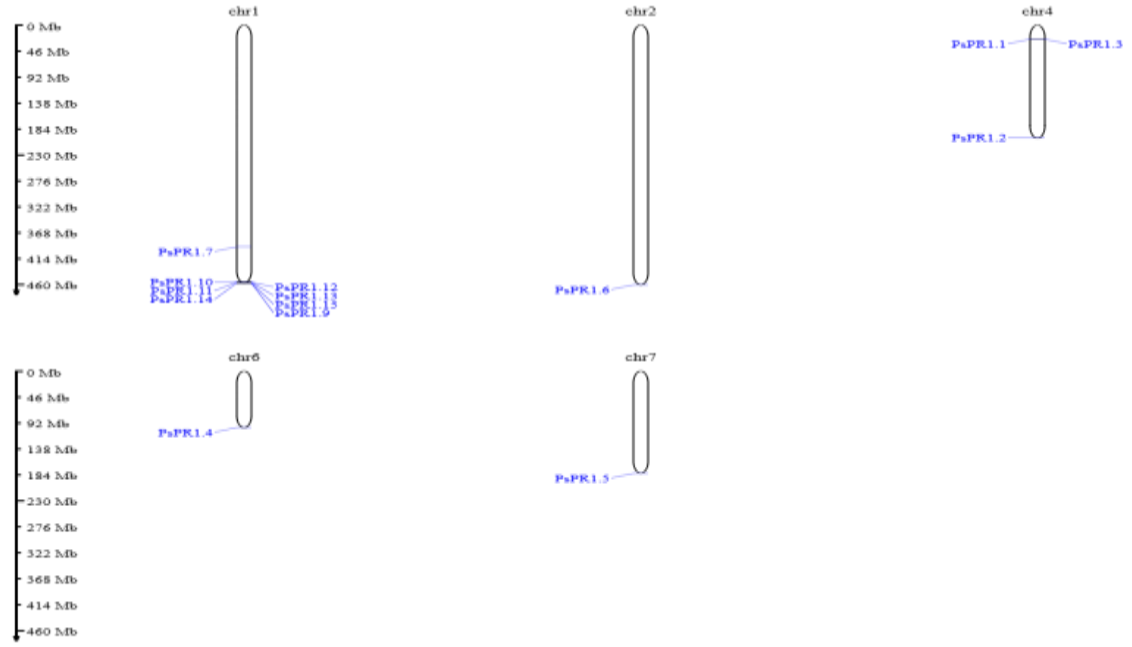

Figure 8: Gene mapping of PR1 genes on the chromosomes of *Pisum sativum*

Synteny and Duplication Patterns

Gene and genome duplications play a key role in evolution by providing the raw genetic material from which new traits can emerge. Consequently, comparative synteny analysis serves as a powerful tool to trace the evolutionary history of genes and genomes, and, ultimately, the traits and organisms

they influence (Zhao and Schranz, 2017). The presence of duplicated PsPR1 gene pairs was confirmed by circular synteny analysis, and these were largely localized within chromosome 1. The gene cluster arrangement on chromosome 1 (as discussed earlier in the chromosomal localization of PR1 genes section) indicates tandem duplication as a major driver of gene family expansion. Additional inter-chromosomal duplication in chromosomes 4 and 7 supports the occurrence of segmental duplication in shaping the current genomic organization of the PsPR1 family.

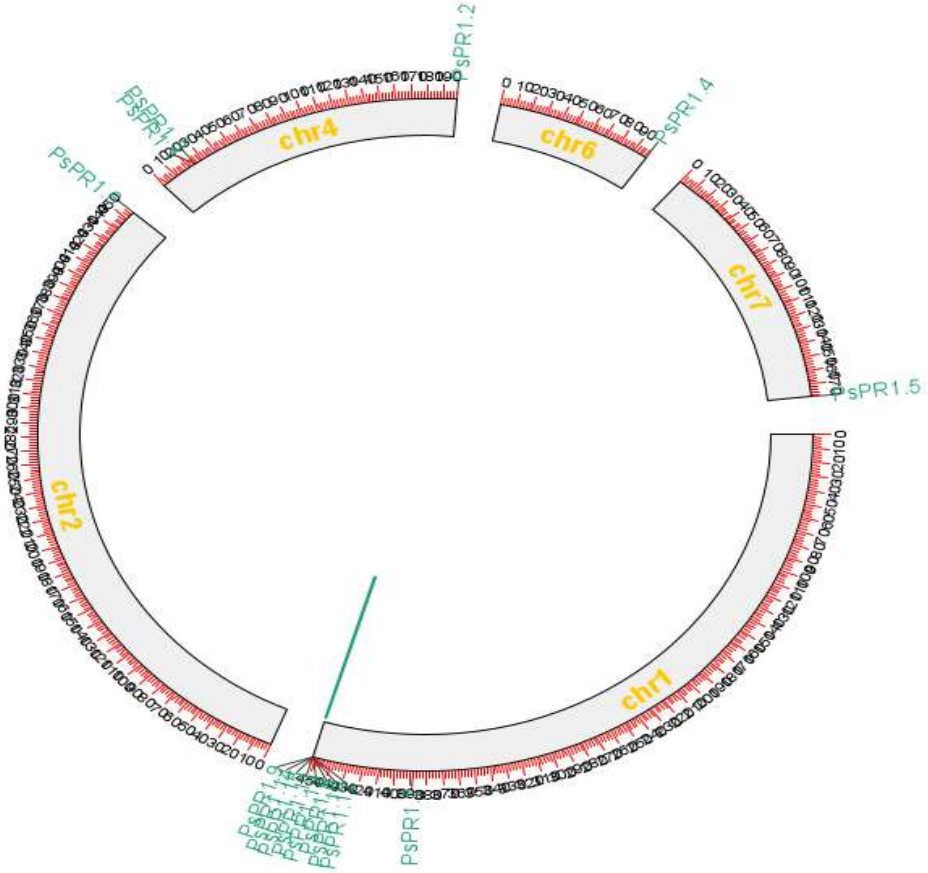

Figure 9: Synteny analysis of PR1 genes in *Pisum sativum*

Conclusion

This study has deeply and across the genome described the PR-1 gene family in *Pisum sativum*. Based on the existence of the conserved CAP domain, fifteen PsPR1 genes were found and verified as legitimate members of the PR-1 family. The majority of genes are intronless or have a single intron, which supports their quick transcriptional activation under stress, according to structural and motif analyses that showed a high degree of conservation. While the occurrence of various intracellular localizations suggests functional diversification, the majority of extracellular localization further supports the established protective role of PR-1 proteins. Tandem duplication, especially on chromosome 1, was a primary factor in the growth of the PsPR1 gene family, with additional contributions from segmental duplication events, according to phylogenetic, chromosomal localization, and synteny analyses. Strong purifying selection is shown by all duplicated gene pairs having Ka/Ks ratios less than one, indicating the evolutionary conservation and functional importance of PR-1 genes in pea defensive mechanisms. Numerous cis-regulatory regions are linked to hormone signalling. Promoter analysis revealed transcription factor binding as well as biotic and abiotic stress responses, underscoring the intricate regulatory networks governing PsPR1 gene expression. PsPR1.1, PsPR1.3, and PsPR1.9 were found to be significantly responsive genes by gene expression study under heat stress conditions; PsPR1.9 exhibited the greatest induction, especially in the heat-tolerant cultivar Akra Chaitra. Together, our results imply that PsPR1 genes are crucial for both heat stress tolerance and pathogen defence. In addition to offering possible candidate genes for creating stress-tolerant pea cultivars via molecular breeding techniques, this study gives a useful basis for further functional characterization.

References

Ali, S., Mishra, A. K., Singh, M., Almalki, M., Friedan, M., Khalifa, A., Chellappan, B. V., Mahmoudi, H., Tyagi, A., and Mir, Z. A. (2025). Genome-wide identification and characterization of pathogenesis related protein 1 gene family in Brassica juncea. *International*

Journal of Biological Macromolecules, 144066. https://doi.org/10.1016/j.ijbiomac.2025.144066

Anuradha, C., Chandrasekar, A., Backiyarani, S., Thangavelu, R., Giribabu, P., and Uma, S. (2022). Genome-wide analysis of pathogenesis-related protein 1 (PR-1) gene family from Musa spp. and its role in defense response during stresses. *Gene* **821**, 146334. https://doi.org/10.1016/j.gene.2022.146334

Brinkman, F. S., and Leipe, D. D. (2001). Phylogenetic analysis. *Bioinformatics: a practical guide to the analysis of genes and proteins* **2**, 349.

Dahl, W. J., Foster, L. M., and Tyler, R. T. (2012). Review of the health benefits of peas (*Pisum sativum* L.). *British Journal of Nutrition* **108**, S3-S10. https://doi.org/10.1017/S0007114512000852

Fan, H., Zhou, J., Wang, Q., Zhang, M., Huang, Z., Han, J., Ying, Y., Yu, Z., and Kai, G. (2025). Comprehensive Genome-Wide Analysis and Expression Profiling of Pathogenesis-Related Protein 1 (PR-1) Genes in Salvia miltiorrhiza. *Food Science & Nutrition* **13**, e70117. https://doi.org/10.1002/fsn3.70117.

Han, Z., Xiong, D., Schneiter, R., and Tian, C. (2023). The function of plant PR1 and other members of the CAP protein superfamily in plant–pathogen interactions. *Molecular plant pathology* **24**, 651-668. https://doi.org/10.1111/mpp.13320

Liu, Y., Engelman, D. M., and Gerstein, M. (2002). Genomic analysis of membrane protein families: abundance and conserved motifs. *Genome biology* **3**, research0054. 1. https://doi.org/10.1186/gb-2002-3-10-research0054

Pavlenko, A., and Pikunova, A. (2024). Mapping of Genetic Factors That Determine Economically Important Traits of Pear (Pyrus) and Marker-Assisted Selection Techniques. *Russian Journal of Genetics* **60**, 551-571. https://doi.org/10.1134/s1022795424700042

Pečenková, T., Pleskot, R., and Žárský, V. (2017). Subcellular localization of Arabidopsis pathogenesis-related 1 (PR1) protein. *International Journal of Molecular Sciences* **18**, 825. https://doi.org/10.3390/ijms18040825

Preissl, S., Gaulton, K. J., and Ren, B. (2023). Characterizing cis-regulatory elements using single-cell epigenomics. *Nature Reviews Genetics* **24**, 21-43. https://doi.org/10.1038/s41576-022-00509-1

Ren, S., Yang, G., He, Y., Wang, Y., Li, Y., and Chen, Z. (2008). The conservation pattern of short linear motifs is highly correlated with the function of interacting protein domains. *BMC genomics* **9**, 452. https://doi.org/10.1186/1471-2164-9-452

Schneiter, R., and Di Pietro, A. (2013). The CAP protein superfamily: function in sterol export and fungal virulence. *Biomolecular concepts* **4**, 519-525. https://doi.org/10.1515/bmc-2013-0021

Smýkal, P., Aubert, G., Burstin, J., Coyne, C. J., Ellis, N. T., Flavell, A. J., Ford, R., Hýbl, M., Macas, J., and Neumann, P. (2012). Pea (*Pisum sativum* L.) in the genomic era. *Agronomy* **2**, 74-115.https://doi.org/10.3390/agronomy2020074

Wang, L., Lu, H., Zhan, J., Shang, Q., Wang, L., Yin, W., Sa, W., and Liang, J. (2022). Pathogenesis-related protein-4 (PR-4) gene family in Qingke (Hordeum vulgare L. var. nudum): genome-wide identification, structural analysis and expression profile under stresses. *Molecular Biology Reports* **49**, 9397-9408. https://doi.org/10.1007/s11033-022-07794-3

Wu, D.-T., Li, W.-X., Wan, J.-J., Hu, Y.-C., Gan, R.-Y., and Zou, L. (2023). A comprehensive review of pea (*Pisum sativum* L.): chemical composition, processing, health benefits, and food applications. *Foods* **12**, 2527. 10.3390/foods12132527

Wufuerjiang, A., Sai, J., Wen, Y., Wang, L., Chen, C., Li, X., and Tian, J. (2025). Genome-Wide Identification of the PR-1 Gene Family in Pyrus betulaefolia Bunge and Its Expression Analysis Under Fire Blight Stress. *International Journal of Molecular Sciences* **26**, 5074. https://doi.org/10.3390/ijms26115074

Yu, C. S., Chen, Y. C., Lu, C. H., and Hwang, J. K. (2006). Prediction of protein subcellular localization. *Proteins: Structure, Function, and Bioinformatics* **64**, 643-651.

Zhao, T., and Schranz, M. E. (2017). Network approaches for plant phylogenomic synteny analysis. *Current opinion in plant biology* **36**, 129-134. https://doi.org/10.1016/j.pbi.2017.03.001

Statements and Declarations

Data Availability Statement

All data are fully available and can be found within the manuscript file.

Acknowledgement

Not Applicable

Conflicts of Interest

The authors declare no conflict of interest.

Ethics Approval Statement

Not applicable.

Ethics Approval Statement

Not applicable.

Consent to Participate

Not applicable.

Consent to Publish

Not applicable.

Author Contributions

Sherazi S.H.U.H. conceived and designed the study, conducted all analyses, interpreted the results, prepared the figures and tables, and drafted and

finalized the manuscript. All co-authors Shafiq, M., Ahmed, H., Ahmad, J., Hammad, M., Sanam, A., Rehan, M., Tufail, M.T., Hayyat, Q. contributed through critical review, intellectual input, and approval of the final manuscript.

Funding
Not applicable.