Open Access

Phytopathogenomics and Disease Control*., *2024, 3(1):1-12
ISSN (Online):2957-5842; ISSN (Print):2957-5834
DOI: https://doi.org/10.22194/Pdc/3.1017
https://societyfia.org/journal/PDC

Society for Innovative Agriculture

Genome-wide identification and characterization of the plant defensins (*Pdfs*) gene family in selected Leguminous crops and their expression profiles in response to biotic and abiotic stresses

Zarqa Hassan[1], Amjad Abbas[1], Adil Zahoor[1,2], Ikhlas Shafique[1], Muhammad Jabran[3], Saman Shahzadi[1], Kinza Ahsan[1], Nafeesa Aman[4], Muhammad Burhan[5] and Muhammad Amjad Ali[1]*

[1]Department of Plant Pathology, University of Agriculture, 38000, Faisalabad, Pakistan; [2]Department of Biotechnology, Chonnam National University, Yeosu 59626, Korea; [3]State Key Laboratory for Biology of Plant Diseases, Institute of Plant Protection, CAAS 100193, Beijing, China; [4]Department of Structural and Environmental Engineering, University of Agriculture, 38000, Faisalabad, Pakistan; [5]Plant Pathology Research Institute, AARI, Faisalabad, Pakistan
*Corresponding author's e-mail: amjad.ali@uaf.edu.pk

Plant defensins (*Pdfs*) represent a subset of cationic antimicrobial peptides characterized by a distinct tertiary structure. Similar to their counterparts in insects and mammals, they play a pivotal role in the innate immunity of plants, exhibiting broad-spectrum antifungal activity. This study focused on a comprehensive genome-wide survey to identify *Pdfs* genes within eight distinct legume crops. This study revealed the presence of Pdfs genes in eight leguminous plants such as 13 in *P. sativum*, 6 in *P. vulgaris*, 16 in *M. trunculata* and 4 in *C. cajon*, 6 in *G. max*, 5 in *A. hypogea*, 5 in *C. arietinum*, and 4 in *V. angularis*. Utilizing various bioinformatic analyses including phylogeny, motif, and gene structure analysis, identified *Pdfs* genes were categorized into two distinct groups i.e., group A (comprising subgroups A1, A2, A3, A4, A5) and group B 2 (B1 and B2). This clustering was based on protein sequences, where the presence of a proline-rich domain at the N-terminal and a linker DNA domain at the C-terminal facilitated differentiation. Gene structural analysis uncovered a variable range of 1 to 3 introns within *Pdfs* genes. Moreover, promoter analysis highlighted the abundance of cis-regulatory elements upstream of *Pdfs* genes, responsive to a diverse spectrum of biotic and abiotic stress stimuli. Expression analysis underlined the crucial role of *Pdfs* protein in defense responses in leguminous plants against both pathogenic and environmental stressors. These findings underscore the involvement of *Pdfs* in strengthening the defense mechanisms of legumes and provide fundamental insights essential for prospective investigations into *Pdfs*.

Keywords: Plant defensins (*Pdfs*); Genome- wide identifications; Gene expressions; phylogenetic analysis; biotic stresses; abiotic stresses; Leguminous Plants.

INTRODUCTION

Leguminosae is an important flowering plant family consisting of 650 to 750 genera and more than 18,000 species of climbers, herbaceous plants, shrubs, and trees. Legumes are very useful to humans and animals as well as agricultural and agroforestry soils. According to Chakraborty *et al.*, (2003). Legumes are one of the most significant crops of the world, with significant propositions for agriculture, animal, environment, and human nutrition, as well as health. Legumes may form symbiotic associations with soil-borne bacteria called rhizobia, which help the plant in fixing atmospheric nitrogen and defend it from fungal infections (Dixon *et al.*, 2003). As a result, they are a major provider of nitrogen to plants across the world. More than 33% of dietary protein comes from grain legumes pea (*Pisum sativum*), lentil (*Lens culinaris*), and common bean (*Phaseolus vulgaris*). Legumes (grain legumes or pulses) are growing in all tropical and sub-tropical areas. Asia contributes 49.2% in pulse production globally (Yanng *et al*., 2000). In Pakistan, its production is 2.6 and 0.4 million tons (Kharif pulses) (Rabi pulses) respectively, which are far less than demand. Abiotic stresses such as drought, heat, cold, frost, water logging, salt, as well as mineral toxicities, reduce the legumes resistance and yield.

Hassan, Z., A. Abbas, A. Zahoor, I. Shafique, M. Jabran, S. Shahzadi, K. Ahsan, N. Aman, M. Burhan and M. Amjad Ali. Genome-wide identification and characterization of the plant defensins (*Pdfs*) gene family in selected Leguminous crops and their expression profiles in response to biotic and abiotic stresses. Phytopathogenomics and Disease Control 3:1-12.
[Received 2 Dec 2023; Accepted 20Mar 2024; Published 4Apr 2024]

Attribution 4.0 International (CC BY 4.0)

Crops that show abiotic stress are more sensitive to insects, weeds, and diseases, resulting in a significant loss of yield (Reddy *et al*., 2004). Biotic and abiotic stresses activate huge varieties of plant responses, changes in gene expression and cellular metabolism lead to changes in growth and crop production. Plants trigger their innate immune system when pathogens attack to combat them (Almasia *et al*., 2008). Innate immunity is an ancient defense system present in most living organisms, including animals, plants, insects, and fungi, and involves nonspecific mechanisms of protection against pathogens (Barocol *et al*., 2002). In plants, innate immunity comprises several mechanisms such as systems of fortification of the cell wall, hypersensitive response, and production of antimicrobial substances such as phytoalexins and pathogenesis-related proteins (Selsted *et al*., 2004). Key members of the class of pathogenesis-related proteins are defensins, cysteine-rich peptides (CRP) with antifungal and antibacterial properties, plus additional functions in development (Mygind *et al*., 2005; Lay and Anderson 2005). *Pdf*s are tiny cysteine-rich peptides that normally consist of an amino acid group, N and C terminal signal peptides that consist of an adjustable region of the cysteine-rich domain (Silverstein *et al*., 2007). Plant defensins control a series of biological actions (Lay *et al*., 2014). Protein synthesis, ion channel function and enzyme activity have been recognized that inhibited the pathogen attack. Some plant defensins have been demonstrated to destroy bacterial growth, but their antifungal effect has been extensively researched (Lay *et al*., 2014). Thomma *et al*., (2002) explained that antimicrobial proteins are the main element of plants' innate defensive response. These proteins include thionine, hevein, knottin-type proteins, lipid transfer proteins, defensins, and snakins.

The propose of this study is to described the bioinformatics analysis and expression profile of the gene family in leguminous plants, as well as the expression pattern of genes in *P. sativum* (pea vegetable), *P. Vulgaris* (common bean plant), *M. trunculata* (barrel medic), *C. Cajon* (pigeon pea), *G. max* (soybean plant*), A. hypogea* (peanut legume), *C. arietinum* (chickpea legume) and *V. angularis* (black-eyed pea), in response to pathogen attack. Comparative phylogenetic analysis, digital expression profiling, homology modeling, and docking analysis were used to learn more about their possible functions during pathogen infections. In this study, we also looked at all the *Pdfs* genes in eight different leguminous plants (Motamayor *et al*., 2013).

Phylogenetic analysis suggested that *Pdfs* proteins from those eight leguminous plants could be divided into five groups. Furthermore, the number of motifs in those *Pdf*s proteins ranged from three to eight, indicating the functional difference. The functional diversity of *Pdfs* genes in leguminous plants was suggested by their expression patterns. When legumes were infected with a pathogen, the gene expression profile revealed that certain *Pdf* genes were engaged in the pathogen response. This research developed a strong platform for further research on *Pdf's* genes in leguminous plants. Genome assembly provides understanding into genome evolution (Hu *et al*., 2015). Our research supports a stable platform for multiple functional analyses of *Pdfs* genes in leguminous plants, as well as the selection of suitable genes for future study. This study presents a complete summary of the *Pdfs* family, setting the stage for future research into the role of *Pdf* in regulating leguminous plant development and stress responses.

MATERIALS AND METHODS

***Data retrieval and identification of Pdfs genes from legume crop species*:** Firstly, we collected 14 reported genes (At1g05830 to At2g02140) encoding plant defensin peptides from *Arabidopsis thaliana* were collected and their sequences were retrieved from TAIR online website (https://www.arabidopsis.org/), Ensembl plant database, Phytozome database, Legume information system(LIS) (https://legumeinfo.org/genomes/gbrowse) respectively (Berardini *et al*., 2015). *A. thaliana Pdf's* genes were used as query sequences to identify *Pdf*'s genes in the genomes of *Arachis hypogea, Cajanus Cajon, Cicer arietinum, Glycine max, Pisum sativum, Phaseolus vulgaris, Medicago truncutula, and Vigna angularis* through RefSeq blast (E-value=$1e^{-4}$) search available at NCBI (https://blast.ncbi.nlm.nih.gov /Blast.cgi). Hereafter, redundant and partial sequences were removed manually. Chromosomal position, genes length, exons, genomic DNA, coding sequences (CDs), and protein sequences were retrieved from NCBI (Wang and Zhang, 2014).

In this study, we used the six letters for the codification of *Pdfs* genes in different species. One letter for genus and two letters for species were used to avoid the overlapping of species under the same genus (Sievers *al*., 2011). One letter for species sources the ambiguity as prefix 'Ah' for *Arachis hypogea*, *Arachis helodes*, and *Arachis hoehnei*. Similarly, code 'Cc' for *Cajanus cajon, Cajanus crassus,Cajanus confertiflorus and Cajanus crassicaulis.* For *Arachis hypogea, Cajanus cajon, Cicer arietinum, Glycine max, Glycine soja, Phaseolus vulgaris, Medicago truncutula, and Vigna angularis, the prefixes, AhyPdf, CcaPdf, CarPdf, PsatPdf, GmaPdf, PvuPdf MtrPdf,* and *VanPd*f were used, respectively.

***Physiochemical properties and subcellular localization of Pdfs protein*:** Predictive modeling provides valuable insights into the behavior of chemicals and materials under various conditions, enabling informed decision-making and risk management across diverse fields of science. To predict the different chemical and physical properties such as isoelectric point (pl), molecular weight (kDa), and many amino acids (aa), help was taken from Sequence Manipulation Suite (Stothard, 2000) having weblink (https://www.bioinformatics.org/sms2/protein_iep.html) and

(https://www.bioinformatics.org/sms/prot_mw.html) respectively. The grand average of hydropathy (GRAVY) of proteins was determined by using weblink (https://www.bioinformatics .org/sms2/protein_gravy.html) (Mukherjee *et al*., 2010). All the data were recorded in table1. To predict the subcellular position, Plant-mPLoc (http://www.csbio. sjtu.edu.cn/bioinf/plant-multi/#) was used (Chou *et al*., 2010). For this purpose, protein sequences of all species were inserted one after another into Plant-mPLoc.

***Multiple Sequence Alignment (MSA) and Phylogenetic Analysis*:** Multiple Sequence Alignment (MSA) is a bioinformatics technique used to compare and align multiple sequences of biological molecules, typically DNA, RNA, or protein sequences. The primary purpose of MSA is to identify regions of similarity and difference among the sequences, allowing researchers to gain insights into evolutionary relationships, functional domains, and conserved motifs. For prediction of evolutionary relationship, MSA was carried out by using the online tool Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) (Madeira *et al*., 2019). Truncated proteins or sequences with no conserved plant defensin domain were excluded manually. Obtained MSA from clustal omega were then introduced in MEGA (version 7.0.21) software (Kumar *et al.,* 2018). For the selection of the best substitution model, the option 'best DNA/protein model' under the heading 'models' in MEGA was selected (Kaplan *et al*., 2001). Phylogenetic trees of all species were built to maximum likelihood with 1000 bootstrap replicates and JTT model. All positions having less than 90% site coverage were removed from MSA. There was total of 238 positions in the final dataset.

***Chromosomal positioning analysis*:** Chromosomal positioning refers to the specific location of a gene or DNA sequence along a chromosome. It is a fundamental aspect of genome organization and function, influencing gene expression, chromosome structure, genome stability, inheritance patterns, and evolutionary dynamics. The species have titled the codes based on chromosomal positioning. The genes that belonged to the *Pdf*A group were named first while those that belonged to the *Pdf*B group were named later in ascending manners. As in *A. hypogea*, *Pdf*A genes are present on chromosomes 3,6,10,12,13,16 and 20 so the coding was started from chromosome 3 while *Pdf*Bis present on chromosome 1,6,11,16 so their coding was started from chromosome 1. Some genes in different species were present on unplaced scaffolds. The genes that clustered in the *Pdf*A group in a phylogenetic tree but were present on an unplaced scaffold were named first. Thereafter, the same method was used for *Pdf*B (Kato *et al*., 2003)

Conserved Motif Analysis: Pfam-like domains having accession no pfam 00304 were confirmed at NCBI Conserved domain database (CDD) (https://www.ncbi.nlm.nih.gov/Structure /cdd/wrpsb.cgi) and their positions were recorded. Pfam-like domains play diverse and essential roles in protein structure, function, and evolution. For graphical visualization of peptidase C-14 domain, NCBI cobalt (https://www.ncbi.nlm.nih.gov/tools/cobalt/re_cobalt.cgi) was used with default parameters (Jones *et al*., 1992). Conserved motifs were determined by MEME database (Bailey, 2009) having weblink (http://meme-suite.org/tools/meme) with following adjusted parameters: No of motif= 25, minimum width= 19, maximum width= 300, minimum site=4, maximum site= 600.

***Gene structure analysis*:** Investigating gene structure is essential for deciphering the organization, regulation, and function of genes in the genome. It provides insights into gene annotation, function prediction, alternative splicing, regulatory elements, evolutionary conservation, and disease mechanisms. For gene structure investigation, CDs and genomics sequences of selected legumes were introduced in Gene Structure Display Server 2.0 (GSDS) (http://gsds.cbi.pku.edu.cn/) (Ling *et al.,* 2020). The introns/exons and UTR were obtained along with the scale.

Cis-regulatory element analysis: From the upstream region of genomic DNA of Arabidopsis and selected legumes, 1kb promotor sequences were retrieved from NCBI. For analysis of cis-elements, these sequences were introduced to the Plant CARE database (http://bioinformatics.psb.ugent.be/webtools/plantcare/ html/) by choosing the 'search for CARE' option from the menu (Lescot *et al*., 2002). Despite the basic cis-regulatory elements (TATA-box and CAAT box), the remaining cis-regulatory elements were recorded. After this, data were further investigated by using TB tools (Toolbox for biologists) v1.046 to construct a heat map of cis-regulatory elements (Chen *et al*., 2020).

Gene-Expression Analysis: The expression profile of *Pdfs* genes from *M. truncatula* in different tissues, involving root, nodule, blade, bud, a pod, and flower tissues were downloaded from the *Medicago truncatulta* Atlas Project (MtGEA, https:// 492 Biochemical Genetics (2019) 57:487–506 1 3 mtgea.noble.org/v3/. Genome-wide transcriptome data from *M. trunculata* under numerous stresses, involving cold, freezing, drought, salt, and bacteria nematode fungi were downloaded from the NCBI short read archive database (SRA database) (https://www.ncbi.nlm.nih.gov) (Accession numbers: SRR10058814-SRR12418415 for abiotic stress and SRR12418419- ERR1421634 for biotic stress) and gene expression was examined using European Galaxy server using weblink(usegalaxy.eu).Heatmaps were drawn with Tbtool (Metsalu*et al*., 2015).

***Protein-Protein Interaction Network Analysis*:** Protein-Protein interaction analysis of these selected legumes protein was performed with the help of an online STRING database using weblink (https://string-db.org/) (Wu *et al*., 2021) Amino acid the sequence of A. *hypogea*, C. *cajon*, C. *arietinum*, *G. max*, *P. vulgaris*, *M. truncutula*, *V. angularis*,

P. sativum protein was used in predicting the interaction network of all these legumes. All parameters of this analysis were used as a default provided on the database. (Zhang *et al*., 2010).

***Molecular Docking of Pdf proteins and Ligands Preparation*:** Five safener structures (44E, ACT, EDO, PA, SO4) were generated in PDB format through ChEBI (Chemical Entities of Biological Interest) database using weblink (http:// www.ebi.ac.uk/chebi/) (Kaplan *et al*., 2001). Pyrx tool was used for the Docking of the *Pdfs* (receptor) with safeners (ligand) using weblink (https://sourceforge.net/projects/pyrx). Discovery studio tool used to build a ligand PDBQT file. (Trott and Olson, 2010). For the docking study, the receptor was run to show ligand-receptor interaction affinity energy. It was used to imagine *Pdfs*-safener binding; discovery studio 2016 software was used. This was also used to visualize hydrogen and hydrophobic ligand-receptor binding interactions (Dallakyan *et al*., 2015).

RESULTS

***Identification of Pdfs gene*:** We identified 75 genes in the genomes of A. thaliana and selected legumes, 13 Pdfs genes in A. thaliana,13 in P. sativum, 6 in P. vulgaris,16 in M. trunculata,4 in C. cajon,6 in G. max,5 in A. hypogea,5 in C. arietinum and 4 in V. angularis. Out of these 73 genes, 49 genes have a theoretical isoelectric value below 9, suggesting that these proteins are acidic while the remaining 27 proteins (above 9) are basic. Also, 51 genes have gDNA <3000 bps, 42 genes have among 3000-8000 bps, 3 genes have >8000 bps. The comprehensive parameters of these genes are listed in supplementary table 1.

Phylogenetic tree and Conserved domain and motif analysis: To conclude the evolutionary history of *Pdfs* genes, a phylogenetic analysis was performed that classified 73 genes into two groups as *PdfA* and *PdfB.* A total of 46 genes belong to *PdfA* while the remaining 27 belong to *pdf*B. *PdfA* genes were subdivided into 5 groups as A1, A2, A3, A4, and A5 (Figure 1). They contain 21,3,12,5, and 5 members, respectively. While *PdfB* genes were subdivided into two groups named B1 and B2. They contain 2 and 25 members, respectively. Subgroup A1, A5, B1, and B2 have all the legumes along with *A. thaliana*. Contrary to previous, subgroup A2, A3, A4 has all legumes but not model plant. Ten different motifs were found during conserved motif analysis by MEME (Figure 1; Table 1). To study the sequence characteristics of the *P. sativum*, *P. vulgaris*, *M. trunculata*, *C. cajon*, *G. max*, *A. hypogea*, *C. arietinum*, and *V. angularis* proteins. Motif 1 comprised the *Pdfs* domain, considered an important motif in these proteins. In addition, the *Pdfs* proteins in the same group obtained similar conserved motifs, which helped the results of the phylogenetic analysis. Multiple sequence alignment was built based on the types of *Pdfs* proteins domains and motifs.

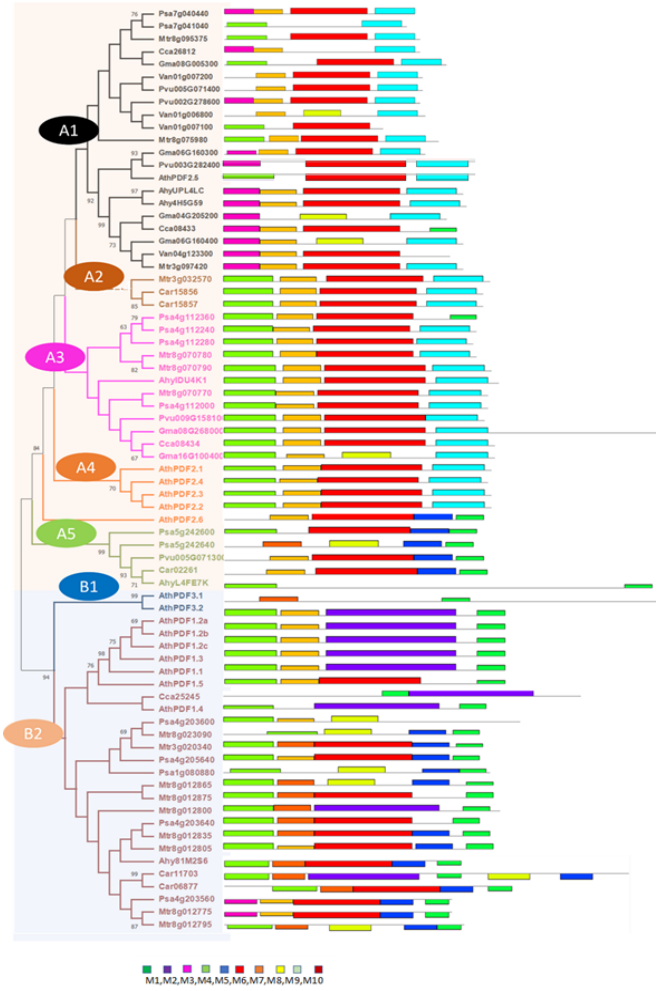

Figure 1. Phylogenetic, conserved motif, analyses of *Pdf* gene family. The phylogenetic tree is classified into two types as *PdfA* and *PdfB*. These types are displayed in different background colors. Subgroups in each type are shown by different alphabets (as for *PdfA* as A1, A2, A3, A4, A5, while for *PdfB* as 'B1, 2B'). Scales show the length of amino acids and genomic size, respectively.

***Gene structure analysis*:** To check the structural diversification of *Pdfs* genes, a gene structure map was built. This map shows the exons, introns, and UTRs distribution patterns in different genes. *PdfA* and *PdfB*seemed to conserve in exon/intron patterns. *PdfA* has a higher number of coding regions e.g., mostly 1 or 2 exons. All genes have 1 to 2 introns with different sizes except AthPDF3.2, It has no introns. A member of subgroup 1A, Mtr2775, has the highest number of introns which are 2 in number. Subgroup 2A, which belongs

Table 1. Different characteristics of the discovered motif using the MEME suite .

No. of motif	Motif sequence	E-value	Sites	Width
1	EARTCESKSHTFKGPCVSDTNCASVCRTE	1.3e-677	50	29
2	GGKCRGFRRRCFCTK	4.0e-281	36	15
3	MARSASLVSTIFVFL	9.0e-180	50	15
4	QKLCEKPSGTWSGVCGNSNACKNQCINLEGAKHGSCN	5.4e-142	9	37
5	LLATLMGPVMV sh	1.2e-117	47	11
6	ACFCYFNC	5.7e-072	34	8
7	EHATFGACHRD	3.3e-043	18	11
8	MEKKRFGFFFL	1.7e-020	14	11
9	LALLLFSTSEV	1.3e-017	12	11
10	CERPSKTFKGPCLS	3.1e-017	11	14

to Mtr2875, shows the greater size of intron which may be due to mutation (e.g., intronic insertion or duplication). All genes are uneven in the presence and size of UTR's. as shown in Figure 2.

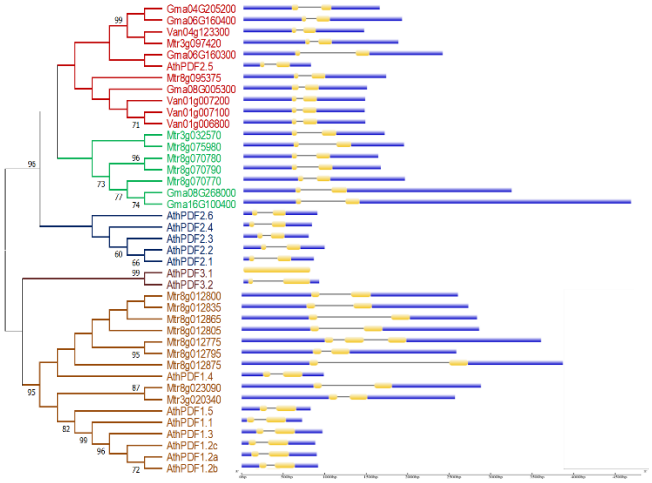

Figure 2. Phylogenetic analysis and gene structure analysis of *Pdf* gene family. The phylogenetic tree is classified into two types as *PdfA* and *PdfB*. These types are displayed in different background colors. Blue and black scales under sections b and c show the length of amino acids and genomic size, respectively.

***Chromosomal positioning*:** To check the distribution of the *Pdfs* gene on these chromosomes, chromosomal maps of different species were arranged by using Map Chart (Figure 3). These disclose the fact that *Pdf's* genes are distributed on different chromosomes rather than evenly distributed. In A. Thaliana *PdfA* genes are present on chromosomes 1,2and 5 Chromosome 2 has many genes that are 4. *C. arietinum* has 5 *Pdf* genes present on 1 and 7 chromosomes. Chromosome 7 has many genes that are 4. In *G. max*, *Pdf* genes are present on chromosomes4,6,8, and 16 Chromosome. The remaining *PdfB* genes are unplaced scaffolds. *M. truncutula* has Pdf's genes distributed on chromosomes 3 and 8. Chromosome 8 has many genes that are an unplaced scaffold. Pdf genes in P. vulgaris are present on chromosomes 2,3,5 and 9. V. angularis Pdf genes are present on chromosomes no 1 and 4. while Cajnus cajon, p. sativum, and Arachis hypogea are unplaced scaffolds.

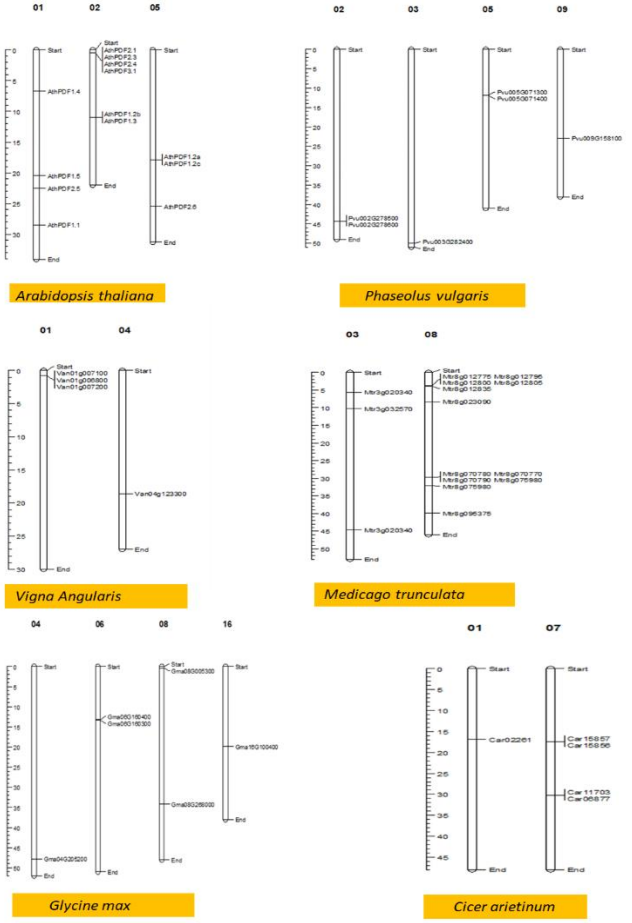

Figure 3. Chromosomal localization of *Pdfs* genes in nine plants. The scale given on the left side of species shows ten million base pairs. On the top, digits show the number of chromosomes containing pdf genes in each species.

Table 2. Different characteristics of a cis-regulatory element under biotic and biotic stress.

Sr.	Site name	Sequence	Function	Category of cis-regulatory element
1	ARE	AAACCA	a cis-acting regulatory element that is essential for the anaerobic induction	Biotic and abiotic stress-responsive element
2	G-box	CACGAC	cis-acting regulatory element involved in the light responsiveness	
3	LTR	CCGAAA	cis-acting element involved in the low-temperature responsiveness	
4	RY-element	CATGCATG	cis-acting regulatory element involved in the seed-specific regulation	
5	ABRE	ACGTG	cis-acting element involved in the abscisic acid responsiveness	
6	TATC-box	TATCCCA	cis-acting element participating in gibberellin-responsiveness	
7	TGACG-motif	TGACG	cis-acting regulatory element involved in the MeJA-responsiveness	
8	CGTCA-motif	CGTCA	cis-acting regulatory element involved in the MeJA-responsiveness	
9	GCN4_motif	TGAGTCA	cis-regulatory element involved in endosperm expression	
10	NON-box	AGATCGACG	cis-acting regulatory element related to meristem specific activation	
11	TC-rich repeats	GTTTTCTTAC	cis-acting element involved in defense and stress responsiveness	
12	CAT-box	GCCACT	cis-acting regulatory element related to meristem expression	
13	AuxRR-core	GGTCCAT	cis-acting regulatory element involved in auxin responsiveness	
14	O2-site	GTTGACGTGA	cis-acting regulatory element involved in zein metabolism regulation	
15	ACE	CTAACGTATT	cis-acting element involved in light responsiveness	
16	TCA-element	CCATCTTTTT	cis-acting element involved in salicylic acid responsiveness	
17	Box II	AAACCA	a cis-acting regulatory element essential for the anaerobic induction	
18	SARE	TTCGACCATCTT	cis-acting element involved in salicylic acid responsiveness	
19	circadian	TCTTAC	cis-acting regulatory element involved in circadian control	
20	A-box	CCGTCC	cis-acting regulatory element	
20	GT1-MOTIF	GGTTAA	light-responsive element	
21	I-box	GATAAGGGT	part of a light-responsive element	
22	LS7	CAGATTTATTTTTA	part of a light-responsive element	Light responsive elements
23	LAMP-element	CTTTATCA	part of a light-responsive element	
24	TCT-motif	TCTTAC	part of a light-responsive element	
25	AE-box	AGAAACTT	part of a module for light response	
26	3-AF1 binding site	TAAGAGAGGAA	light-responsive element	
27	AT1-motif	AATTATTTTTTATT	part of a light-responsive module	
28	GATA-motif	AAGGATAAGG	part of a light-responsive element	
29	CAG-motif	GAAAGGCAGAC	part of a light response element	
30	AAAC-motif	CAATCAAAACCT	light-responsive element	
31	CHS-CMA1a	TTACTTAA	part of a light-responsive element	
32	CHS-CMA2a	TCACTTGA	part of a light-responsive element	
33	sbp-CMA1c	CTTTATCTCTTCCA	part of a light-responsive element	
34	chs-Unit 1 m1	ACCTAACCCGC	part of a light-responsive element	
35	LAMP-element	CTTTATCA	part of a light-responsive element	
36	MRE	AACCTAA	MYB binding site involved in light responsiveness.	
37	CCAAT-box	CAACGG	MYBHv1 binding site	
38	MBSI	aaaAaaC(G/C) GTTA	MYBHv1 binding site	
39	MBS	CAACTG	MYBHv1 binding site	
40	Box 4	ATTAAT	part of a conserved DNA module involved in light responsiveness	Development related element
41	3-AF3 binding site	CACTATCTAAC	module array (CMA3)	
42	P-box	CCTTTTG	gibberellin-responsive element	
43	GARE-motif	TCTGTTG	Gibberellin-responsive element	
44	TGA-element	AACGAC	auxin-responsive element	
45	GC-motif	CCCCCG	enhancer-like element involved in anoxic specific inducibility	
46	HD-Zip 1	CAAT(A/T) ATTG	element involved in differentiation of the palisade mesophyll cells	
47	HD-Zip 3	GTAAT(G/C) ATTAC	protein binding site	

***Cis-regulatory element promoter analysis*:** To study the cis element in promoter regions of *Pdfs* gene located in 2000 bp from the transcriptional start site of upstream region were predicted by the PLANT CARE database. Through these observations there were 47 different kinds of response elements, such as light responsive element, metabolism regulation element, defense and stress responsive element involved in drought, salt, low temperature and anaerobic, and hormone responsive element associated with salicylic pathways were found (Figure 4 and Table 2). Similarly, the

defense and stress responsive elements were found in the promoter region of 27 *Pdfs* genes, in which 47 contained development related response elements. According to the function of these genes, different numbers of cis-regulatory elements (CREs) are present at several positions in the promotor region. CREs are linear sequences of non-coding DNA that permit the transcriptional factor to bind. In this way, they regulate the expression or movement of genes. Promotor analysis exposes the function of different genes in different species when exposed to either biotic/abiotic stresses or developmental changes. There are three major groups of cis-regulatory elements. These are hormonal, stress, and development related elements. Many cis-regulatory elements are involved in biotic or abiotic stress's reaction. Among these, light responsiveness elements are positioned at the first position in the number. After this, cis elements that contribute to hormonal responsiveness are distinguished. Development sensitive elements come at third position.

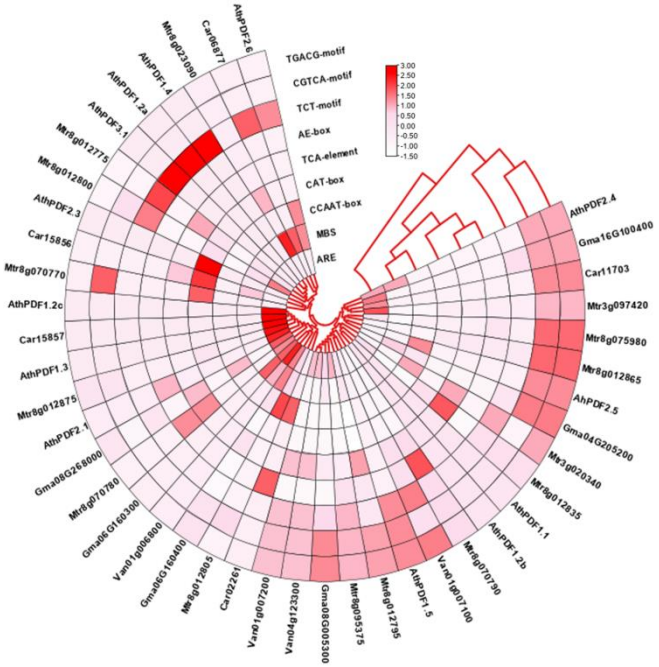

Figure 4. This heatmap illustrated the existence or absence of promotors on an upstream region of a gene. Scale on the rightest side shows several conserved motif or boxes in the 1kb sequence of promotor region. This scale varied according to the number of regulatory elements. Red and blue colors for large and no cis-regulatory elements.

***Protein-protein interaction (PPI) network analysis*:** Protein-protein interaction (PPI) analysis was performed with the help of a string online server to predict the interaction of *Pdf* resistance protein with other proteins working inside of the cell. Analysis showed that this Pdf protein interacts with several proteins of the cell. It was expected that *PDF2.1* resistance proteincooperateswith transcriptional factors of the R genes, for instance DAA1 protein (Fig. 5).

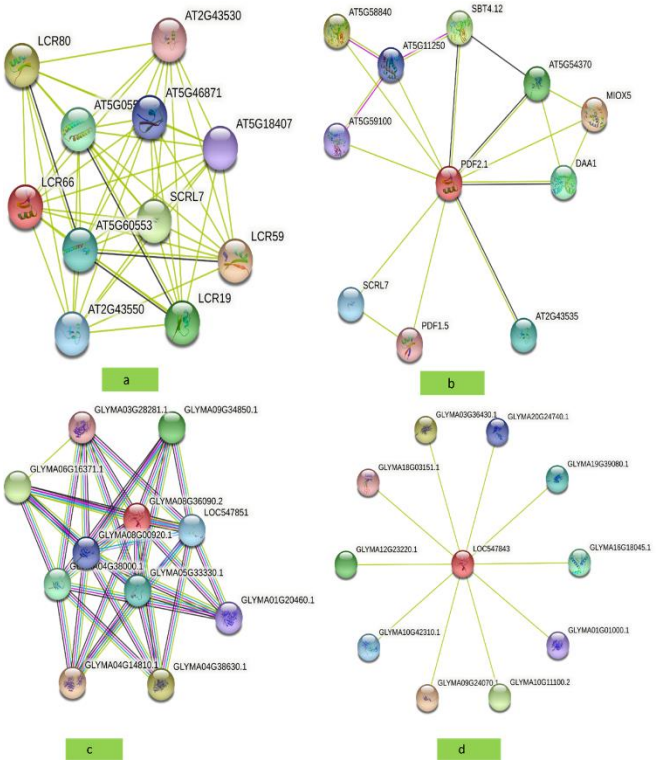

Figure 5. Protein-Protein interaction analysis of *Pdfs* protein. Nodes of the network represent different proteins that interact with the query protein.

Query protein is shown in the center of the network through crimson color. Colored but filled nodes show that the corresponding protein is of known 3-D structure and shows the first level of interaction with query protein. Colored but empty nodes show that the corresponding protein is of unknown 3-D structure, however, represents the first level of interaction. The scale provided below shows the nature of reference in predicting the above network.

***Expression profiles of Pdfs genes under biotic and abiotic stress*:** To examine the expression profile of *M. truncatula* genes in response to biotic (fungi, nematode, and bacteria) and abiotic stresses (cold,drought, and salt), RNA seq data were retrieved from the NCBI SRA database (Table 3). This data was further investigated and the heatmap was constructed to represent the expression level of *Pdf*s genes. Based on the expressional profiles of Mtr genes, they were divided into two groups (Figure 6). Gene profiling of the differentially expressed gene showed that three genes were downregulated after the salt cold and drought treatment for 2

h and upregulated after 6 h and 12 h. Similarly, four genes of *M. hapla* show upregulation while three genes of *S. medicae* show downregulation.

Table 3. Expression analysis of *Pdfs* gene through SRA database under biotic and abiotic stress.

	SSR/ERR used for abiotic stresses		SSR/ERR used for biotic stresses
1	SRR10058814	1	SRR12418416
2	SRR10058815	2	SRR12418417
3	SRR10058816	3	SRR12418418
4	SRR10058817	4	SRR12418419
5	SRR10058818	5	SRR12418420
6	SRR10058819	6	SRR8797556
7	SRR10058820	7	SRR8797566
8	SRR10058821	8	SRR8797567
9	SRR10058822	9	SRR8797568
10	SRR10058823	10	SRR8797572
11	SRR10058824	11	SRR8797555
12	SRR10058825	12	ERR1421639
13	SRR12418415	13	ERR1421640
		14	ERR1421641
		15	ERR1421642
		16	ERR1421643
		17	ERR1421688
		18	ERR1421689
		19	ERR1421691
		20	ERR1421756
		21	ERR1421634

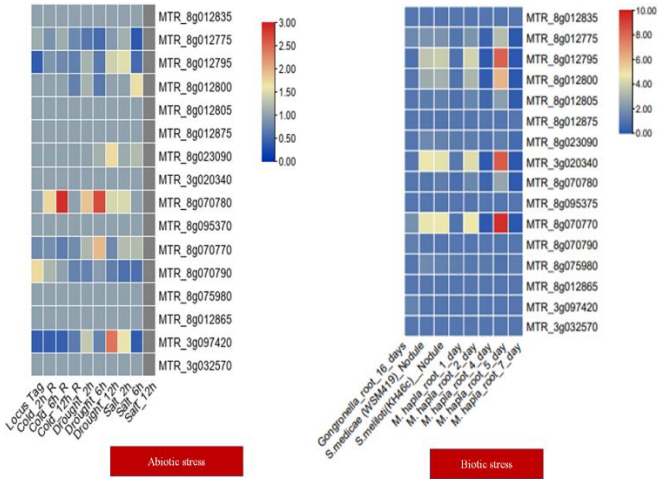

Figure 6. Expression profile of *Pdfs* genes in different tissues during abiotic stresses and different biotic stresses.

***Molecular Docking analysis*:** Five ligands (44E, ACT, EDO, PA, SO4) were retrieved through Chembll database (Table 4). With the original protein that is 2lr3 protein, these ligands were docked, and their score is compared to the original Ligand Docked score. These types of ligands are selected and are compared with those ligands whose energy Is minimum. This was done to attain stability between protein-ligand interaction. Out of these ligands, the 44E PubChem compound has a score of -4.8 kcalmol, and hence 44E ligand is considered the best Ligand. The Image reveals original Ligand interacting residues within the active site regions of 2lr3 protein. In this way, eight H-bonds formed by the Ligand with the active residues of 2lr3 (Figure 7).

Table 4. Complete list of ligands with binding affinity from database.

Ligand	Binding affinity	rmsd/ub	rmsd/lb
2lr3_model1_44E_uff_E=361.45	-4.8	0.000	0.000
2lr3_model1_44E_uff_E=361.45	-4.7	6.198	2.154
2lr3_model1_44E_uff_E=361.45	-4.7	3.926	2.122
2lr3_model1_44E_uff_E=361.45	-4.7	4.899	3.385
2lr3_model1_44E_uff_E=361.45	-4.7	3.453	2.109
2lr3_model1_44E_uff_E=361.45	-4.6	2.755	1.958
2lr3_model1_44E_uff_E=361.45	-4.6	6.124	3.481
2lr3_model1_44E_uff_E=361.45	-4.5	2.843	2.075
2lr3_model1_44E_uff_E=361.45	-4.5	4.794	1.967
2lr3_model1_ACT_uff_E=0.00	-2.3	0.000	0.000
2lr3_model1_ACT_uff_E=0.00	-2.2	1.701	1.331
2lr3_model1_ACT_uff_E=0.00	-2.1	12.438	11.909
2lr3_model1_ACT_uff_E=0.00	-2.1	12.962	12.499
2lr3_model1_ACT_uff_E=0.00	-2.1	13.955	13.727
2lr3_model1_ACT_uff_E=0.00	-2.1	2.365	1.749
2lr3_model1_ACT_uff_E=0.00	-2.0	4.611	4.195
2lr3_model1_ACT_uff_E=0.00	-1.9	2.966	2.740
2lr3_model1_ACT_uff_E=0.00	-1.8	2.572	2.260
2lr3_model1_EDO_uff_E=8.12	-2.5	0.000	0.000
2lr3_model1_EDO_uff_E=8.12	-2.4	2.180	0.352
2lr3_model1_EDO_uff_E=8.12	-2.4	11.275	10.886
2lr3_model1_EDO_uff_E=8.12	-2.3	2.241	1.651
2lr3_model1_EDO_uff_E=8.12	-2.3	2.786	1.444
2lr3_model1_EDO_uff_E=8.12	-2.2	14.321	14.117
2lr3_model1_EDO_uff_E=8.12	-2.2	14.259	14.152
2lr3_model1_EDO_uff_E=8.12	-2.1	12.919	12.787
2lr3_model1_EDO_uff_E=8.12	-2.1	13.051	12.867
2lr3_model1_PA_uff_E=340.32	-3.7	0.000	0.000
2lr3_model1_PA_uff_E=340.32	-3.7	9.393	8.405
2lr3_model1_PA_uff_E=340.32	-3.5	9.045	8.253
2lr3_model1_PA_uff_E=340.32	-3.5	16.435	13.991
2lr3_model1_PA_uff_E=340.32	-3.5	16.501	14.163
2lr3_model1_PA_uff_E=340.32	-3.5	16.077	14.230
2lr3_model1_PA_uff_E=340.32	-3.5	15.558	13.450
2lr3_model1_PA_uff_E=340.32	-3.4	9.102	7.898
2lr3_model1_PA_uff_E=340.32	-3.4	9.671	8.478
2lr3_model1_SO4_uff_E=579.79	-2.9	0.000	0.000
2lr3_model1_SO4_uff_E=579.79	-2.8	11.033	10.508
2lr3_model1_SO4_uff_E=579.79	-2.7	15.735	15.012
2lr3_model1_SO4_uff_E=579.79	-2.7	15.465	14.695
2lr3_model1_SO4_uff_E=579.79	-2.6	17.061	16.295
2lr3_model1_SO4_uff_E=579.79	-2.5	14.647	14.077
2lr3_model1_SO4_uff_E=579.79	-2.5	13.532	13.023
2lr3_model1_SO4_uff_E=579.79	-2.5	18.253	17.388
2lr3_model1_SO4_uff_E=579.79	-2.4	13.563	13.390

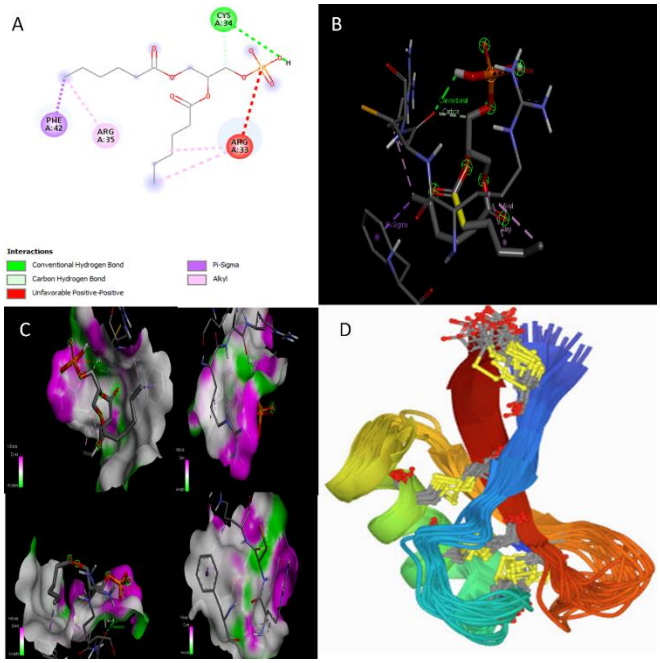

Figure 7. A 2D diagram interactions of protein model. B, 44E ligand interacting within 2lr3 protein of *Medicago trunculata* in active sideregion. C, The Original Ligand showed 4 interactions with O of 2lr3 protein.

Figure 7 shows that model protein is attached with the best ligand at its active sides. This figure shows eight hydrogen bonds are formed when ligand is binding protein at its active side residue. D, Structure of 2lr3 protein model with the interdomain region in red, blue, and in green color. This protein model is obtained through PDB online database which is find there 2lr3 protein model. This structure shows alpha, beta, hexa structure.

DISCUSSION

In this current study, firstly we identified 75 *Pdfs* genes in different legumes plants and analyzed their chemical properties and genomic distributions. The RNAseq data analyses provide the *Pdfs* some conceivable roles in growth development of *Medicago trunculata* under different some abiotic and biotic stress responses. *Pdf* gene family in *Medicago trunculata* plant provide some important functions in the growth, development, and stress responses to plants in higher plants. In several tissues, orthologous gene pairs displayed similar expression patterns. These results suggest that some gene pairs retained similar functions during evolution, while others exhibited functional diversification (Yao *et al*., 2019).

Leguminous plants, a subset of the *Fabaceae* family, are vital sources of essential nutrients and amino acids globally. This botanical family includes diverse species such as *Pisum sativum* (pea), *Phaseolus vulgaris* (common bean), *Medicago truncatula* (barrel medic), *Cajanus cajon* (pigeon pea), *Glycine max* (soybean), *Arachis hypogea* (peanut), *Cicer arietinum* (chickpea), and *Vigna angularis* (black-eyed pea) (Graham *et al*., 2003). These plants face numerous biotic and abiotic stress factors throughout their lifecycles. Their ability to survive and thrive relies heavily on the efficiency and effectiveness of their defense mechanisms (Ahmadi *et al*., 2020). One of the crucial elements of this defense is the Defensin protein family, a group of Pathogen-Related (PR) proteins in class 12. These proteins play multifaceted roles, including the inhibition of phytopathogenic fungi and involvement in biotic stress response mechanisms (Zhu, 2008).

We identified total 75 *Pdf* (Plant defensin) gene family, for instance, *Arabidopsis thaliana* contains 15 *Pdf* genes, while *Medicago truncatula* has 16, *Pisum sativum* has 13, *Phaseolus vulgaris* has 6, *Glycine max* has 6, *Arachis hypogea* has 5, *Cicer arietinum* has 5, and *Vigna angularis* has 4. Notably, some species possess a smaller number of *Pdf* genes compared to model plants (Kim *et al*., 2013; Wang and Zhang, 2014; Liu *et al*., 2016). To one side from Mtr8g012865 (which resides on cell membrane), some different proteins are found in the outer region of cell. In several plants extracellular localization of *Pdf* proteins has also been previously reported. in the plasma membrane, cytoplasm, and the nucleus the localization of *Pdf* proteins has also been briefly described. Some factors such as protein–protein interaction and post-translational modifications are involved in the variations of subcellular localization of plant cells. The 3D structure prediction of these small proteins can provide some valuable information about functions of protein based on ligand-binding sites. In this study, some amino acids such as, proline, cysteine lysine, serine leucine, and threonine were commonly predicted as the key binding residues in the structure of *Pdf* proteins, in which some amino acids such as proline, serine, and leucine are associated with responses to different environmental stimuli (Abdullah *et al*., 2021).

Post-translational modifications are processes of chemical modifications of proteins, and they produce diversity in structure and function, including subcellular localization, protein–protein interaction, and regulating enzyme activity by allosteric phenomena (Nahirñak *et al*., 2016). The mechanism of phosphorylation of proteins also plays an important role in cell signaling, regulation of different mechanisms of proteins, and provides as a substrate for different kinases (Kim *et al*., 2013).

The phylogenetic analyses divided this *Pdf* protein into 2 main groups and 5 subgroups respectively. The motifs analysis and intron-exon analysis display some variations inside proteins which are cluster together in same phylogeny. These results represent that *Pdf* of some groups may have evolved during evolution process which are further involved

in the variations of motifs and introns in some groups. These findings also show that some other processes are related to the function of proteins instead of their close phylogenetic relationships. A similar observation was reported in *Nicotiana tabacum* (Nawaz *et al*., 2019). However, some studies also proposed that the closely related proteins on a phylogenetic tree have a similar function.

Similarly, some cis act as regulatory elements which are involved in transcription of regulation of genes and tempted through some independent signal transduction pathways under various biotic and abiotic stress response. We find various key cis-regulating elements in response to different factors such as light, stresses, hormones, and growth in the promoter region site. Some other cis-regulating elements for anerobic induction, drought, low temperature, and plant defense were also found (Li *et al*., 2022). The presence of varied cis-regulating elements in promoter regions represent some basic roles in the regulation of diverse legumes pathways. This Study found that *Pdf* genes are intricately involved in responses to multiple stresses.

Under conditions of drought, cold, salt stress, bacterial infection, nematode infection, and fungal infection, specific Mtr genes displayed both upregulation and downregulation. Additionally, the presence of cis-elements (LTR) in some genes suggests involvement in low-temperature responsiveness (Ye *et al*., 2019). However, some LTR-containing genes exhibited downregulation during cold and freezing stresses. These findings highlight the dynamic regulatory mechanisms that *Pdf* genes employ in response to a spectrum of biotic and abiotic stressors (Ye *et al*., 2019). *Pdf* proteins can be classified into two types: *PdfA* and *PdfB*. While *PdfA* proteins predominantly function in a basic medium, *PdfB* proteins operate effectively in an acidic environment (Yao *et al*., 2019). These Pdf proteins are localized in various cell compartments, including chloroplasts, nuclei, and the cytoplasm.

The study revealed that despite variations in the number of *Pdf* genes, all selected leguminous species exhibited equal or greater *Pdf* gene counts compared to Arabidopsis thaliana. Phylogenetic analysis showed a close relationship between Athpdf 1.1 and Past genes, clustering them together. Gene structure analysis indicated that Athpdf 3.2 lacked introns. Conserved motif analysis demonstrated close relationships between Athpdf, Car, Psat, and Mtr genes. Similarly, motif analysis revealed identical motif patterns between Car and Mtr genes within the same groups.

In terms of *Pdf* classification, the study identified 46 genes as PdfA and 27 as PdfB. PdfA genes were found to be more prevalent than PdfB genes, aligning with existing literature. Notably, only 12 members exhibited the LSD1 domain at the N-terminus, as determined by multiple sequence alignment (MSA). This discrepancy suggests variations in intronic duplication or insertion, observable in gene structure mapping. PdfA genes typically possess one intact exon, whereas PdfB genes have 1-2 exons. For instance, Mtr8g012775 contained two exons akin to PdfB genes, though it clustered under PdfA in the phylogenetic tree. This observation might be attributed to the absence of peptidase C14 domain-encoding proteins. Both Mtr8g012775 and PdfB genes consist of 136 amino acids, leading to the inclusion of the former in PdfA, challenging the conventional classification. Additionally, a docking analysis revealed prevalent binding modes of the 44E ligand with the known three-dimensional 2lr3 protein mode (Morris *et al*., 2008).

***Conclusion*:** This work delivers some new insight into the characterization, identification, and expressions of the *Pdf* genes in the different eight leguminous plants. This study successfully identified and characterized 75 Plant defensin (*Pdf)* genes within selected leguminous species through a range of bioinformatic approaches. The analysis encompassed phylogenetic clustering, chromosomal localization, gene structure mapping, protein interaction, synteny analysis, conserved domain organization, promoter region investigation, docking analysis, and role of *Pdf* genes in legumes underscores their significance in both plant development and defense responses. RNA seq analyses describe the role of the *Pdf* genes in growth and yield development. This study reveals that *Medicago trunculata* genes are varied because of their structure and other regulatory systems. This analysis also summaries that they are involved in different cellular pathways which are related to different development and stress response actions. This research not only sheds light on the intricate mechanisms within leguminous plants but also provides a foundation for further investigations into *Pdf* genes in diverse and less explored legume species, promising valuable contributions to plant biology and agriculture. Similarly, the result obtained through this research indicates that the *Pdf* genes provides the resistance response against some biotic stress (fungi, bacteria, nematode) and abiotic stress (heat, cold, drought) which causes dramatic yield losses to *production* of legumes each year.

***Authors' contributions*:** Zarqa Hassan authored this paper, Muhammad Amjad Ali and Amjad Abbas reviewed the technical aspects of article. Saman Shahzadi, Kinza Ahsan and Ikhlas Shafique worked on tables. Nafeesa Aman and Muhammad Burhan contributed to literature sources. Adil Zahoor and Muhammad Jabran contributed to the computational formatting of the paper.

***Conflict of interest*:** The authors declare that there is no conflict of interest.

Funding: There is no funding for this paper.

***Ethical statement*:** For this article, no studies with human participants or animals were performed by any of the authors.

All studies mentioned were in accordance with the ethical standards indicated in each case.

***Availability of data and material*:** We declare that the submitted manuscript is our work, which has not been published before and is not currently being considered for publication elsewhere.

***Code availability*:** Not applicable.

***Consent to participate*:** All authors participated in this research study.

***Consent for publication*:** All authors submitted consent to publish this research article in PDC.

REFERENCES

Abdullah, F.S., F. Mehmood, H. M. Malik, I. Ahmed, P. Heidari and P. Poczai. 2021. The GASA gene family in cacao genome wide identification and expression analysis. Agronomy 11:7-1425.

Zadeh, M. A., M. Chen, J. T. Chen, S. Hasanzadeh, S. Ahmar, and P. Heidari. 2020. Insights into the genes involved in the ethylene biosynthesis pathway in Arabidopsis thaliana and Oryza sativa. Journal of Genetic Engineering and Biotech. *18:1* Journal of Genetic Engineering and Biotech 18:1-20.

Almasia, N.I., A.A. Bazzini, H. Hopp, C. Vazquez-Rovere. 2008. Overexpression of snakin-1 gene enhances resistance to *Rhizoctonia solani* and *Erwinia carotovora* in transgenic potato plants. Molecular Plant Pathology 9:329-338.

Bailey, T.L., M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble. 2009. MEME Suite: tools for motif discovery and searching. Nucleic acids Research 1:37(suppl_2): W202-8.

Berardini, T.Z., L. Reiser, D. Li, Y. Mezheritsky, R. Muller, E. Strait, and E. Huala. 2015. The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. Genesis 53:474-485.

Chakraborty, U., B. Sarkar, and B.N. Chakraborty. 2003. Protection of soybean rot by *Bradyrhizobium japonicum* and *Trichoderma harzianum* associated with changes in enzyme activities and phytoalexin production. Journal of Mycology and Plant Pathology 33:21-25.

Chen, C., H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He, and R. Xia. 2020. TBtools: an integrative toolkit developed for interactive analyses of big biological data. Molecular Plant 13:1194-1202.

Chou, K.C. and H.B. Shen. 2010. Plant-mPLoc: a top-down strategy to augment the powerfor predicting plant protein subcellular localization. PLoS One. 5:1-11.

Dallakyan, S. and A.J. Olson. 2015. Small-molecule library screening by docking with PyRx. In Chemical biology. Humana Press, New York, NY. Pp. 243-250.

Dixon, R.A. and L.W. Sumner. 2003. Legume natural products: understanding and manipulating Jones complex pathways for human and animal health. Plant Physiology 131:878-885.

Graham, P.H. and Vance, P. Carroll. 2003. Legumes: Importance and constraints to greater use. Plant Physiology 131:872-7.

Hu, B., J. Jin, A. Y. Guo, H. Zhang, J. Luo, and G. Gao. 2015. GSDS 2.0: an upgraded gene features visualization server. Bioinformatics 31:1296-1297.

Jones, D. T., W. R. Taylor and J.M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8:275-282.

Kaplan, W. and T.G. Littlejohn. 2001. Swiss-PDB viewer (deep view). Briefings in Bioinformatics 2:195-197.

Kato, T., S. Sato, Y. Nakamura, T. Kaneko, E. Asamizu, and S. Tabata, 2003. Structural analysis of a *Lotus japonic* genome. V. Sequence features and mapping of sixty- four TAC clones which cover the 6.4 Mb regions of the genome. DNA Research 10:277- 285.

Kim, S.M., C. Bae, S.K. Oh, and D. Choi. 2013. A pepper (*Capsicum annuum* L.) meta caspase 9 (Camc9) plays a role in pathogen-induced cell death in plants. Molecular Plant Pathology 14:557-566.

Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura.2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35:1547-1554.

Lay, F.T., S. Poon, J. A. McKenna, A. A. Connelly, B. L. Barbeta, B. S. McGinness, J. L. Fox, N. L. Daly, D. J. Craik, R. L. Heath, and M. A. Anderson. 2014. The C-terminal propeptide of a plant defensin confers cytoprotective and subcellular targeting functions. BMC Plant Biology 14:1-13.

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

Li, Y., X. Liu, Y. Xiao, Y. Wen, K. Li, Z. Ma, L. Yang, Y. Zhu, and J. Yin. 2022. Genome-wide characterization and function analysis uncovered roles of wheat LIMs in responding to adverse stresses and TaLIM8-4D function as a susceptible gene. Plant Genome 15:3- e20246.

Ling, L., Y. Qu, J. Zhu, D. Wang, and C. Guo. 2020. Genome-wide identification and expression analysis of the VQ gene family in Cicer arietinum and Medicago truncatula. Peer J. *8*: e8471.

Liu H, J. Liu, Y. Wei 2016. Identification and analysis of the metacaspase gene family in tomatoes. Biochemical and biophysical research communications 479:523-529.

Madeira, F., Y. M. Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A. R. Tivey, S. C. Potter, R. D. Finn, and R. Lopez. 2019. The EMBL-EBI search and sequence analysis tools APIs. Nucleic Acids Research 47:636-641.

Metsalu, T. and J. Vilo. 2015. ClustVis: a web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. Nucleic Acids Research 43:566- 570.

Morris, G.M. and Lim-Wilby. 2008. Molecular docking. In Molecular modeling of proteins. Humana Press.pp. 365-382.

Motamayor, J.C., K. Mockaitis, J. Schmutz, N. Haiminen, D. L. Iii, O. Cornejo, S. D. Findley, P. Zheng, F. Utro, S. Royaert, and C. Saski. 2013. The genome sequence of the most widely cultivated cacao type and its use to identify candidate genes regulating pod color. Genome biology 14:1-25.

Mukherjee, S., A. Szilagyi, A. Roy and Y. Zhang. 2010. Genome-wide protein structure prediction in Multiscale Approaches to Protein Modeling. Springer 11:255-279.

Mygind, P.H., R. L. Fischer, K. M. Schnorr, M. T. Hansen, C. P. Sönksen, S. Ludvigsen, D. Raventós, S. Buskov, B. Christensen, L. De Maria, and O. Taboureau. 2005. Plectasin is a peptide antibiotic with therapeutic potential from a saprophytic fungus. Nature 437:975-80.

Nahirñak, V., M. Rivarola, M. Gonzalez de Urreta, N. Paniego, H. E. Hopp, N. I. Almasia, and C. Vazquez-Rovere. 2016. Genome-wide analysis of the Snakin/GASA gene family in *Solanum tuberosum* cv. Kennebec. American journal of Potato Research 93:172-188.

Nawaz, Z., K. U. Kakar, R. Ullah, S. Yu, J. Zhang, Q. Y. Shu, and X. L. Ren. 2019. Genome-wide identification, evolution and expression analysis of cyclic nucleotide-gated channels in tobacco (*Nicotiana tabacum* L.). Genomics 111:142-158.

Reddy, A.R., K.V. Chaitanya, and M. Vivekanandan. 2004. Droughtinduced responses of photosynthesis and antioxidant metabolism in higher plants. Journal of Plant Physiology 161:1189-1202.

Selsted, M. E. 2004. Theta-defensins: cyclic antimicrobial peptides produced by binaryligation of truncated alpha-defensins. Current Protein and Peptide Science 5:365-367.

Sievers, F., A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, and J. D. Thompson. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology 7- 539.

Silverstein, K.A., Moskal Jr, W.A., Wu, H.C., Underw. 2007. Small cysteine rich peptides resembling antimicrobial peptides have been under-predicted in plants. The Plant Journal 51:262-280.

Stothard, P. 2000. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 28:1102-1104.

Thomma, B.P., Cammue, B.P. and Thevissen, K 2002. Plant defensins. Planta 216:193- s202.

Trott, O. and A. J. Olson. 2010. Auto Dock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of computational chemistry 31:455-461.

Wang, L. and H. Zhang. 2014. Genomewide survey and characterization of metacaspase gene family in rice (*Oryza sativa*). Journal of genetics 93:93-102.

Wu, L., Y. Chen, M. Chen, Y. Yang, Z. Che, Q. Li, X. You, and W. Fu. 2021. Application of network pharmacology and molecular docking to elucidate the potential mechanism of Astragalus–Scorpion against prostate cancer. Andrologi*a* 53:9-e14165.

Yao, J., J. S. Luo, Y. Xiao and Z. Zhang. 2019. The plant defensin gene AtPDF2. 1 mediates ammonium metabolism by regulating glutamine synthetase activity in *Arabidopsis thaliana*. BMC Plant Biology 19:1-13.

Ye, C., Q. Zhou, X. Wu, G. Ji, and Q. Q. Li. 2019. Genome-wide alternative polyadenylation dynamics in response to biotic and abiotic stresses in rice. Ecotoxicology and environmental safety 183:109-485.

Zhang, Y., P. Gao and J. S. Yuan. 2010. Plant protein-protein interaction network and interactome. Current genomics 11:40-46.

Zhu, S. 2008. Discovery of six families of fungal defensin-like peptides provides insights into origin and evolution of the CSab defensins. Molecular immunology 45: 828-38.