Int.J.Curr.Microbiol.App.Sci (2020) Special Issue-11: 305-315

International Journal of Current Microbiology and Applied Sciences
ISSN: 2319-7706 **Special Issue-11 pp. 305-315**
Journal homepage: http://www.ijcmas.com

Original Research Article

In-silico and Bioinformatic Analysis of Pathogenesis Related Protein-1 in *Cicer arietinum*

Meenakshi Verma*

Sardar Vallabhbhai Patel University of Agriculture and Technology,
Meerut (Uttar Pradesh), India

**Corresponding author*

A B S T R A C T

Objective of the study is *in-silico* analysis of pathogenesis related (PR-1) proteins in *Cicer arietinum.* Hypothesis: The pathogenesis-related protein 1 (PR-1) gene family plays important roles in the plant metabolism in response to biotic and abiotic stresses. The present study aimed genome-wide identification and bioinformatics analyses of PR-1 genes in chickpea (*Cicer arietinum*). protein sequences of chickpea pathogenesis related protein-1 (PR-1) were obtained from NCBI protein reference database. These candidate proteins were comprehensively analysed with, multiple sequence alignment (CLUSTALW), domain architecture studies (CDD), signal peptide and motif extraction (MEME) followed by phylogenetic analysis using MEGA software. Other properties like molecular weight, average PI, number of amino acids and Open Reading Frame (ORF) length was calculated using Protein- isoelectric point calculator was obtained to see properties of PR-1 proteins. We analysed 11 pathogenesis related protein-1 (PR-1) genes/proteins that were obtained by blastp of tomato proteins in Chickpea genome with the help of bioedit. The presence of PF00188 and SM000198 domain structures from NCBI Conserved Domain databases confirmed they are PR-1 proteins. All the 11 proteins of *Cicer arietinum* have molecular weight in given range of PR-1 proteins. In view of the p^I data, it is understood that al1 of the 13 PR-1 proteins are acidic. Six PR-1 genes were in chromosome number 1 (Ca1). Using MEME online program tool, a total of 15 distinct motifs named 1-15 were detected, only motif 1 (red), motif 2 (cyan) and motif 3 (green) were found to be associated with the CAP domain (PF00188). These three motifs (motif 1, 2 and 3) were found in all PR-1 proteins as expected. These genes can be explored for their utilization in the future molecular biology applications in crop improvement programmes. Also, PR-1 genes can be good candidates as molecular markers for developing varieties resistant/tolerant to abiotic stresses.

Keywords

Pathogenesis-related protein 1, *Cicer arietinum*, Plant metabolism, Bioinformatics analyses, CAP domain

Introduction

Plants have developed complex mechanisms to protect themselves against pathogens. Pathogenesis-related proteins are those which are produced as a result of plant–pathogen interactions during pathogenicity and are cysteine-rich, low molecular weight, and anti-microbial in their mode of action. They regulate production of several proteins, peptides or compounds which are toxic to pathogens or prevent pathogen infections where they start [1]. Previously, the PR proteins were divided into 5 families based on their localization, isoelectric point, molecular mass and biological activity, while,

currently they are categorized into 17 families. The PR proteins of the same family are known to have highly homologous sequences and similar functions. Having antifungal activities, PR-1 constitutes the main family of the PR proteins induced by pathogens or salicylic acid [3]. The first member of PR-1 family, PR1-a was identified in Nicotiana tabacum plants infected with tobacco mosaic virus (TMV) [4]. Homologues of tobacco PR1 proteins have been identified in many plant species such as Arabidopsis thaliana (At), barley, tomato, maize, rice, etc. 4. The PR-1 family belongs to PR proteins (known as antimicrobial peptides, AMPs) which are classified into 17 families based on their protein sequence similarities, enzymatic activities and other biological features [2]. These proteins can also be defined as acidic or alkaline based on their theoretical isoelectric point, cysteine-rich secretory protein, antigen 5 and pathogenesis-related-1 (CAP) proteins [6]. A CAP domain contains approximately 150 amino acids, and CAP-containing proteins are observed in more than 2500 species including plants, animals, bacteria and fungi with diverse roles [7].

Materials and Methods

Experimental site

The experiment was conducted at Indian Grassland and Fodder Research Institute (IGFRI), Jhansi, Uttar Pradesh.

Gene identification procedures in chickpea genome

First, protein sequences of tomato PR-1 reference such as Solyc00g174340, Solyc01g106600, Solyc01g106610, Solyc01g106620, Solyc01g106640, Solyc02g065470, Solyc07g006700, Solyc07g006710, Solyc08g068990, Solyc09g006010 and Solyc09g007010 was obtained from (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_ Slycopersicum). Later, these sequences were used for blastp analyses in chickpea genome annotation (ITAG2.4) (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_ Carietinum) with threshold as e-30 [19]. Further, to analyse the conservation of domains in PR1 proteins (query and predicted), batch Web CD-search Tool was used. Molecular weight, average P^I, number of amino acids and Open Reading Frame (ORF) length was calculated using Protein-isoelectric point calculator ("http://isoelectric.org/") was obtained to see properties of PR-1 proteins.

Phylogenetic analysis

The unrooted phylogenetic tree of identified PR-1 genes/proteins domains in *Cicer arietinum* was constructed using MEGA (Molecular Evolutionary Genetic Analysis) ver.X. The reliability of each tree was established by conducting 1000 bootstrap sampling steps.

Multiple sequence alignment

Multiple sequence alignment of amino acid sequences of PR-1 genes/proteins in *Cicer arietinum* was performed using CLUSTALW. It is a general-purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences.

Motif analysis

The online tool Multiple Expectation Maximization for Motif Elication (MEME SUITE) 5.0.1 (http://memesuite.org/tools/meme) was utilized for identifying the conserved motifs with the optimized

parameter as follows: distribution of any motifs- any occurrence, maximum number of motifs-15, minimum sites-2, maximum sites-600 and minimum width-6.

Results and Discussions

We analysed 11 pathogenesis related protein-1 (PR-1) genes/proteins that were obtained by blastp of tomato proteins in Chickpea genome with the help of ("https://bioedit.software.informer.com"). Similar result was observed by (Akbudak *et al.,* 2020) he discovered 13 novel SlPR-1 genes, each of which produces a protein belonging to the CAP superfamily (PF00188). (Tellis *et al.,* 2017) also reported 92 candidate PR1 proteins through the PSI-BLAST and HMMER programs. Mitshuhara *et al.,* (2008) reported 12 PR-1 genes using rice genome database. The presence of PF00188 and SM000198 domain structures from Pfam 32.0 ("http://pfam.xfam.org") (Finn *et al.,* 2016) and SMART ("http://smart.embl-heidelberg.de/") (Letunic *et al.,* 2018) databases were investigated for the confirmation of the obtained PR-1 proteins. SMART or NCBI (CDD) is a web resource for the identification and annotation of protein domains and the analysis of protein domain architectures. Outcome of this analysis is the presence of conserved domains in their sequences, which indicate that they are homologous. All the 11 proteins of *Cicer arietinum* has molecular weight in range 15-25Kda, (Loon *et al*., 1999) also mentioned in his study that PR protein are low-molecular weight proteins (6-43 kDa). In view of the p^I data, it is understood that al1 of the 13 PR-1 proteins are acidic. SignalP 4.1 server 9 was used to predict the presence and location of signal peptide cleavage sites in amino acid sequences. Six PR-1 genes were in chromosome number 1 (Ca1). Maximum genes were present in Ca1.

Two phylogenetic trees were constructed one was constructed to determine phylogeny relationship within 11 protein sequences of *Cicer arietinum*, phylogenetic tree was divided into two main groups as shown in figure 1. Group 1 consists seven protein sequences whereas group 2 consists three protein sequences, these groups decides closeness between the different gene IDs. Other phylogenetic tree was constructed to determine phylogeny relationship among PR-1 genes/proteins from four different plants viz., *Cicer arietinum, Solanum lycopersicum, Medicago truncatula and Arabidopsis thaliana* to analyse how closely they are related to each other. According to figure 2 *Cicer arietinum* is closely related to all the three plants mentioned.

The conserved motifs of PR-1 proteins were analysed to explore the similarity and diversity of motif composition. A total of 15 distinct motifs named 1-15 were detected as shown in (Fig. 3). Only motif 1 (red), motif 2 (cyan) and motif 3 (green) were found to be associated with the CAP domain (PF00188). These three motifs (motif 1, 2 and 3) were found in all PR-1 proteins as expected. Remarkably, motif 10 (SFSLKC) was found only in Ca_00197 and Ca_03269 protein sequences. The lowest number of motifs was identified as five in Ca_16022 protein sequences. These variations of the protein motifs in number, types and positions may be related with molecular function and roles in different chickpea metabolic pathways. Similar results were presented by (Akbudak *et al.,* 2020) that total 10 conserved motifs were identified using MEME, and only motif 1 (red) and motif 2 (cyan) were found to be associated with the CAP domain (PF00188) also, (Tellis *et al.,* 2017) reported that only two motifs could be annotated using pfam database. Motif 1 (29 residues) was annotated as CAP domain and motif 9 (200 residues) was annotated as Pkinase Tyr domain (Fig. 4; Table 1 and 2).

Table.1 Characteristics of the PR-1 genes/proteins in *Cicer arietinum*

Gene ID	mRNA ID	Chromosome/Scaffold	Position	Molecular wt. (kDa)	P^I	AA length	ORF length	Signal peptides
CaPR1_1	Ca00197	Ca1	1607638-1608129	18.3	7.96	164	492	1-23
CaPR1_2	Ca00199	Ca1	1617816-1618259	16.53	7.49	148	444	-
CaPR1_3	Ca00200	Ca1	1625456-1625899	16.5	7.49	148	444	-
CaPR1_4	Ca00196	Ca1	1604432-1604869	15.75	5.25	146	438	-
CaPR1_5	Ca14762	Ca1	17749171-17749725	20.48	7.71	185	555	1-27
CaPR1_6	Ca00195	Ca1	1592472-1592909	16.28	4.82	146	438	-
CaPR1_7	Ca03269	Ca7	2041598-2042125	19.9	7.4	176	528	1-27
CaPR1_8	Ca03270	Ca7	2039155-2039763	23.81	5.7	203	609	1-23
CaPR1_9	Ca02098	Ca8	4592962-4592501	19.66	5.83	180	540	1-26
CaPR1_10	Ca16022	Ca7	13212266-13212724	17.03	7.75	153	459	-
CaPR1_11	Ca14029	Ca4	19318679-19319326	25.33	6.57	216	648	1-25

Table.2 Regular expression sequences of 15 motifs

Motif	Sequence	Width	e-value
Motif-1	EC[GL]HYTQ[VI]VW[RK][KD]SLR[ILV]GCAKVKCDNGGTF	29	5.60E-136
Motif-2	D[MW][TS][GP]T[ED]AVKLW[VA]DEKPYYDY[NY][RT]NSCVD	27	3.00E-81
Motif-3	CNY[DS]PPGN[YR][IPV]G[EQ]RPY	15	3.20E-77
Motif-4	[IVL][PG][PDN][LVI][VY]WD[EK][TKS][VL][AE][SA][FY]A[RQ][TNW]YANQRK	22	2.30E-59
Motif-5	H[AL]QNS[PA][QS][DE][YF][VL][KD][AS]HN[KI]AR[FAS][EN]V[GS]	21	5.10E-53
Motif-6	[HS][SGN]N[GN][PR]YGEN[IL][AF][WI][SG][ST]G	15	1.80E-29
Motif-7	PL[ET]WSE[KQ]LA[KN][DT]TS[KL]LVRYQR[DN][KR]M[GS]C[DQ]FANLT[AE][GS]KYG[AG]NQLW	41	8.00E-11
Motif-8	DCQL[KM]H	6	1.10E-08
Motif-9	[ML][KLT][MSV]H[FMT]F[LP][CFL]F	9	4.90E-01
Motif-10	EF[LV][FY][AR]HN[LW]VR	10	2.10E+01
Motif-11	L[PS]LI[IS]I	6	3.40E+02
Motif-12	FLF[LV][LM][ST][FT]T	8	3.70E+02
Motif-13	[KN]DC[AE]LE	6	1.20E+03
Motif-14	[IV]EH[KS][FK]P	6	1.50E+03
Motif-15	S[FM]SL[KL]C	6	2.40E+03

Fig.1 Phylogenetic analysis of 11 PR-1 genes/proteins in *Cicer arietinum*. The phylogenetic tree was generated by the maximum-likelihood (ML) method using MEGA 7.0. The bootstrap values were 1000 replications for major branches

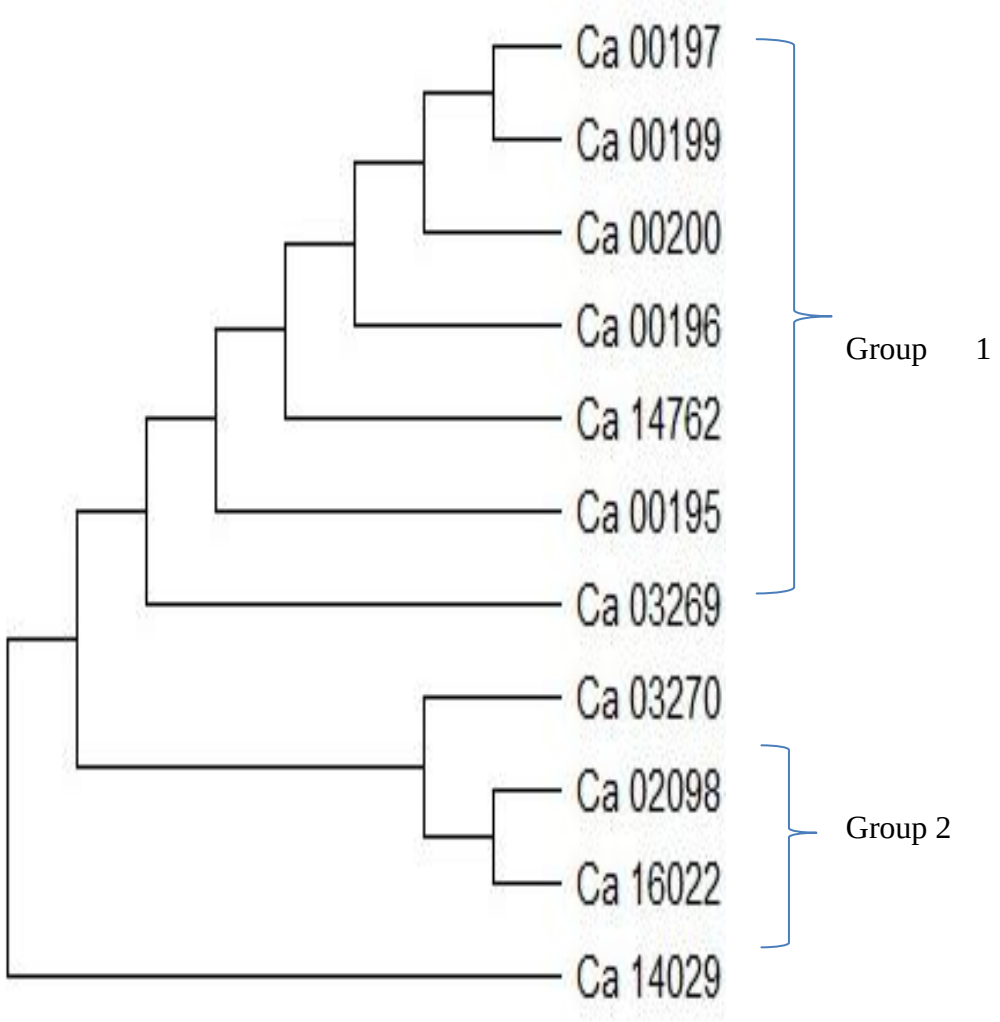


Fig.2 Phylogenetic analysis of PR-1 genes/proteins from *Cicer arietinum, Solanum lycopersicum, Medicago truncatula and Arabidopsis thaliana*

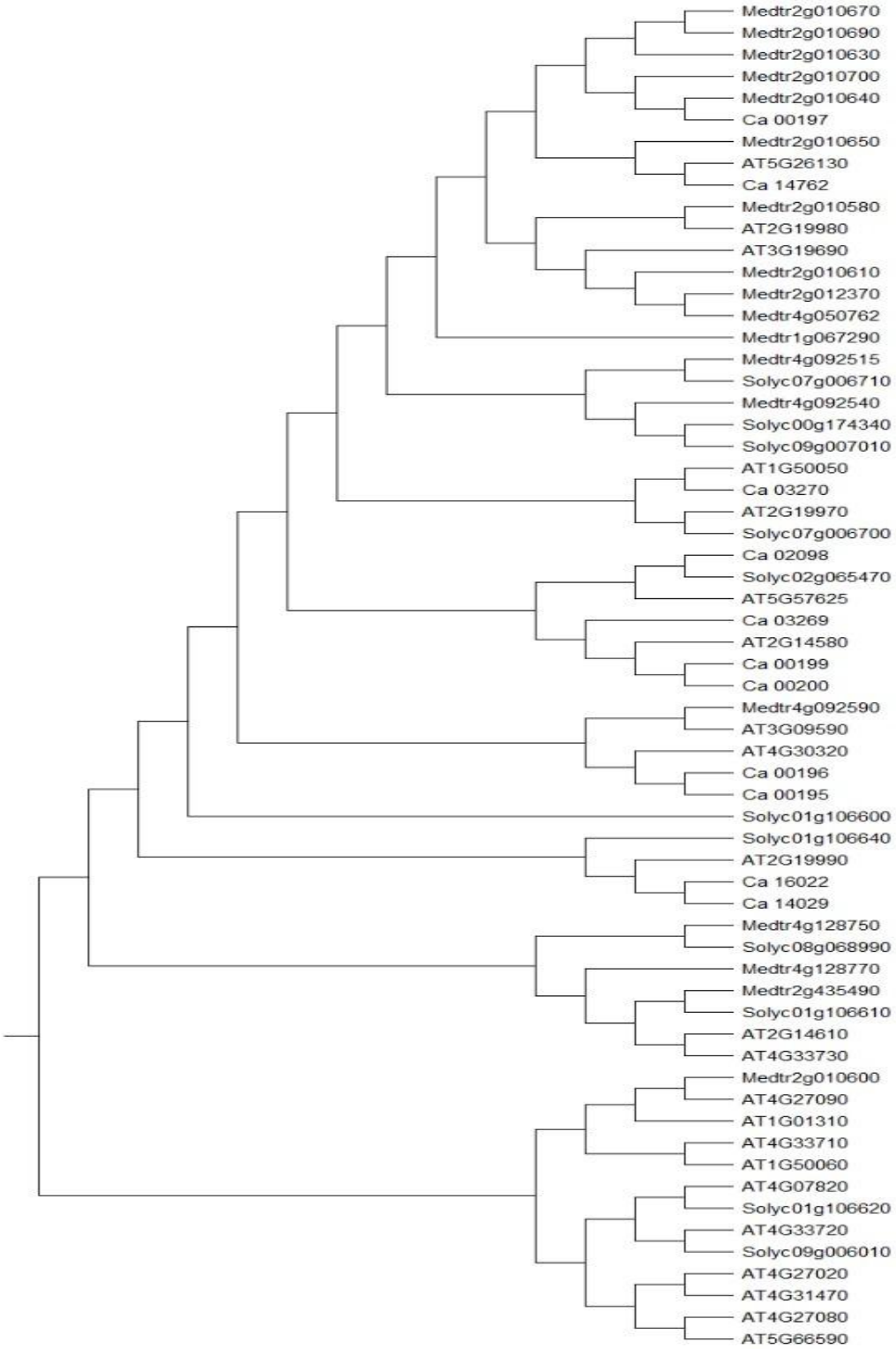


Fig.3 Multiple sequence alignment of PR-1 genes/proteins using the BioEdit 7.2.5 software. The red rectangle indicates the signal peptide sequences found by SignalP-5.0. The blue shows the PF00188 domain structure. pfam00188 is a member of the superfamily ("https://www.ncbi.nlm.nih.gov/Structure/cdd/cl00133")

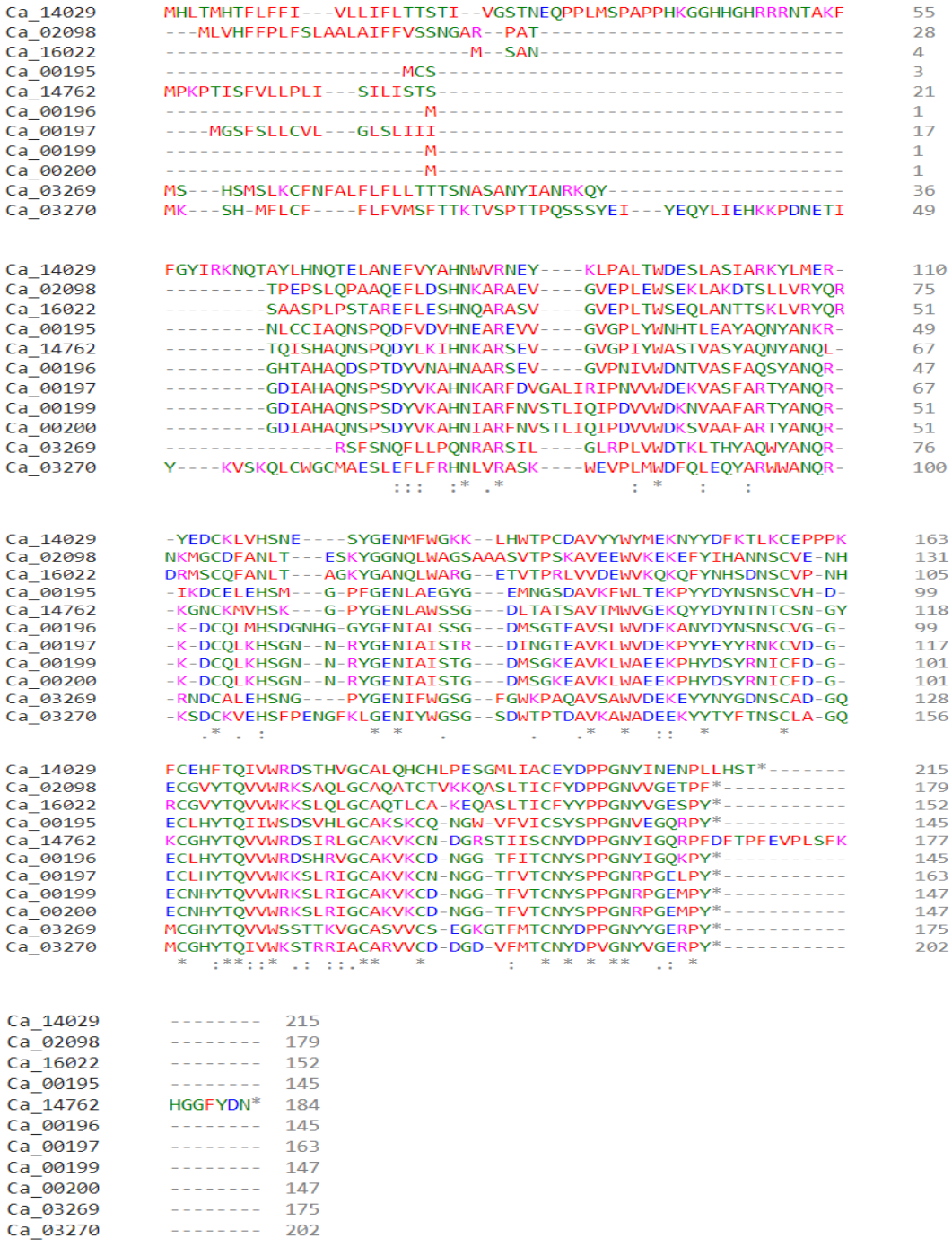
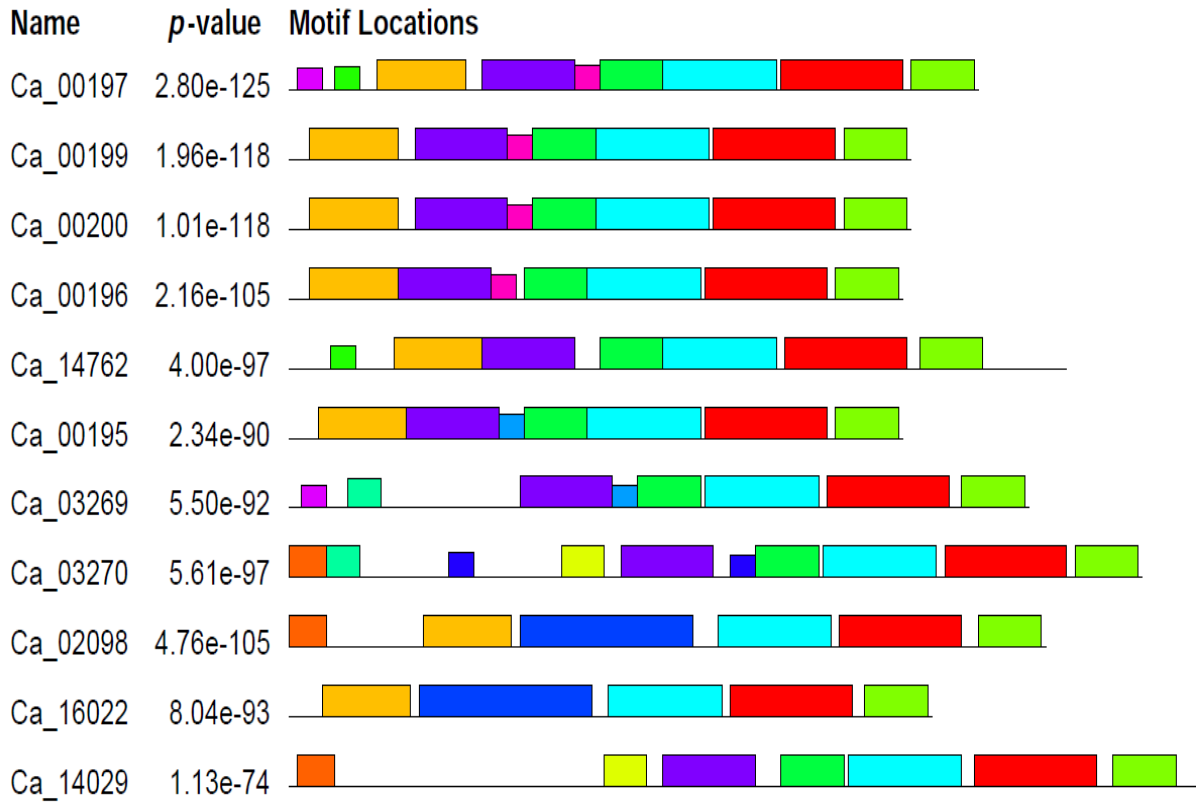


Fig.4 Phylogenetic relationships and organisation of conserved motifs of PR-1 genes/proteins sequences in *Cicer arietinum.* The motifs identified by MEME software are represented by coloured boxes and their consensus sequences are shown in Table 2



Motif	Symbol	Motif Consensus
1.		ECGHYTQVVWRKSLRIGCAKVKCBNGGTF
2.		DMTPTEAVKLWVDEKPYYDYYTNSCVD
3.		CNYDPPGNYVGERPY
4.		IPPLVWDKTVASYARTYANQRK
5.		HAQNSPSDYVKAHNKARFEVG
6.		HSNGPYGENIAWSSG
7.		PLEWSEKLAKDTSKLVRYQRDKMGCDFANLTEGKYGGNQLW
8.		DCQLKH
9.		MKVHFFLFF
10.		EFLFRHNWVR
11.		LPLIII
12.		FLFLLSFT
13.		KDCELE
14.		IEHKFP
15.		SFSLKC

PR-1 gene family play important roles in abiotic stress response as well as biotic stress response in plants. Besides, it can be claimed that the SlPR-1 genes are good candidates as molecular markers for developing varieties resistant/tolerant to abiotic stresses. These PR-1 genes can be explored for their utilization in the future molecular biology applications in crop improvement programmes. PR-1 genes are frequently used as marker gene for systemic acquired resistance in many plant species.

Acknowledgments

I gratefully acknowledge the Indian Grassland and Fodder Research Institute (IGFRI), Jhansi, Uttar Pradesh who permitted to carry out this research work by using all the facilities available at Indian Grassland and Fodder Research Institute (IGFRI).

References

Akbudak, M. A., Yildiz, S., and Filiz, E. 2020. Pathogenesis related protein-1 (PR-1) genes in tomato (*Solanum lycopersicum* L.): Bioinformatics analyses and expression profiles in response to drought stress. Genomics, 112(6): 4089-4099.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., and Salazar, G. A. 2016. The Pfam protein families database: towards a more sustainable future. *Nucleic acids research*, *44*(D1): D279-D285.

Gibbs, G. M., Roelants, K., and O'bryan, M. K. 2008. The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 protein—roles in reproduction, cancer, and immune defense. *Endocrine reviews*, *29*(7): 865-897.

Van Loon, L. and E. Van Strien 1999. The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins, Physiological and molecular plant pathology, 55, 85-97.

van Loon, L.C. and A. van Kammen 1970. Polyacrylamide disc electrophoresis of the soluble leaf proteins from Nicotiana tabacum var. "Samsun" and "Samsun NN". II. Changes in protein constitution after infection with tobacco mosaic virus, Virology, 40, 190-211.

Letunic, I., and Bork, P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic acids research*, *46*(D1): D493-D496.

Mitsuhara, I., Iwai, T., Seo, S., Yanagawa, Y., Kawahigasi, H., Hirose, S., ... and Ohashi, Y. 2008. Characteristic expression of twelve rice PR1 family genes in response to pathogen infection, wounding, and defense-related signal compounds (121/180). Molecular Genetics and Genomics, 279(4): 415-427.

Punja, Z. K. 2001. Genetic engineering of plants to enhance resistance to fungal pathogens—a review of progress and future prospects. *Canadian Journal of Plant Pathology*, *23*(3): 216-235.

Ryals, J., Uknes, S., and Ward, E. 1994. Systemic acquired resistance. *Plant physiology*, *104*(4): 1109.

Schneiter, R., and Di Pietro, A. 2013. The CAP protein superfamily: function in sterol export and fungal virulence. *Biomolecular concepts*, *4*(5): 519-525.

Sels, J., Mathys, J., De Coninck, B. M., Cammue, B. P., and De Bolle, M. F. 2008. Plant pathogenesis-related (PR) proteins: a focus on PR

peptides. *Plant Physiology and Biochemistry*, *46*(11): 941-950.

Tellis, M., Mathur, M., Gurjar, G., Kadoo, N., and Gupta, V. 2017. Identification and functionality prediction of pathogenesis- related protein 1 from legume family. Proteins: Structure, Function, and Bioinformatics, 85(11): 2066-2080.