World Research Journal of Bioinformatics
E-ISSN : 2348-5566, Volume 1, Issue 1, 2013, pp.-001-008.
Available online at http://www.bioinfopublication.org/jouarchive.php?opt=&jouid=BPJ0000298

TOLL AND INTERLEUKIN-1 RECEPTOR (TIR) DOMAINS IN *Cajanus cajan*: AN *IN-SILICO* PERSPECTIVE

SINGH V.K.[1], SINGH A.K.[2] SINGH N.K.[3] AND SINGH B.D.[1]*

[1]School of Biotechnology, Faculty of Science, Banaras Hindu University, Varanasi- 221 005, UP, India.
[2]Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi- 221 005, UP, India.
[3]National Research Center on Plant Biotechnology, Indian Agricultural Research Institute, New Delhi- 110 012, India.
*Corresponding Author: Email- brahmadsingh@gmail.com

Received: November 07, 2013; Accepted: December 03, 2013

Abstract- Toll and interleukin-1 receptor (TIR) domain is associated with plant disease resistance (R) proteins known to play important role in plant immunity by regulating defense responses against invading pathogens. To verify this, we identified the candidate genes associated with TIR signature from pigeonpea (*Cajanus cajan*). Using various *in silico* techniques 103 genes were successfully identified from the available datasets. *Cis*-acting element analysis revealed that the identified genes have hormone responsive elements, seed-specific elements and biotic and abiotic stress specific elements. The hormone response elements included gibberellin, salicylic acid, abscisic acid, MeJA, ethylene and auxin response elements. The Skn-1, GCN4 and RY elements are seed-specific and impart endosperm specific expression. In case of biotic and abiotic stress elements, TC-rich repeats are involved in defense and stress response, WUN-motif is concerned with wound response, HSE and MBS are for heat and drought response, respectively, while Box W1 and AT-rich sequences function as fungal elicitor elements for maximal elicitor-mediated activation. *In silico* expression studies suggest that 5 genes seem to have major roles in seed developmental stages during 10DAF (Days after Flowering) to 42DAF based on pod length and seed size. TIR domain consistency was further verified in 38 cultivars of *C. cajan*. PCR products of TIR domain from one candidate gene (AFSP01036179.1_ 5692_ 6123) amplified from three cultivars BSMR-846, MAL-13 and Bahar were successfully deposited with accession numbers JX667781.1, JX667782.1 and JX667783.1, respectively. Structural analysis of the TIR domain with PMDB number PM0078097 suggested that this domain is mainly involved in signaling and auto-regulation with respect to plant immunity.

Keywords- *In-silico*, *Cajanus cajan*, *R* gene, TIR, Domain

Introduction

Plants, like animals, are also able to launch successful defense responses against invading pathogens. For this, plants have developed a variety of strategies including molecular, chemical, and physical barriers to infection [1]. Plant disease resistance (R) proteins mediate strong immune responses to pathogens. Toll and interleukin-1 receptor (TIR) domains in plants exists as a component of a family of multi-domain proteins known as resistance genes. These R proteins also contain a nucleotide-binding (NB) domain and/or a leucine-rich repeat (LRR) region [2]. They are involved in self-association, signaling, and auto-regulation [3]. Toll-like receptors (TLRs) are a class of proteins that play a key role in the innate immune system. The plant Toll/IL-1 receptor/plant disease resistance gene (TIR) domains play an integral role in the immune system forming part of a defense mechanism against a number of microbial infections. The C-terminal boundary of R protein TIR domains was previously defined to occur at the site of the first intron, based on alignments with animal TIR domain sequences [4]. The TIR structure consists of five β-strands forming a parallel β-sheet at the core of the protein. The β-strands are connected by a series of α-helices and the overall fold mimics closely that of other mammalian and bacterial TIR domains [5]. First plant TIR domain containing protein was reported in tobacco [6].

In view of the above, *in silico* analysis of *C. cajan* genome sequence available in the NCBI database was undertaken to identify TIR-associated genes, and to determine various features of these genes, including sequential and structural classification.

Materials and Methods

In-silico analysis of TIR proteins from *Cajanus cajan*

In silico analysis of the genome sequence of *Arabidopsis thaliana* reveals that this model plant encodes at least 135 proteins containing TIR domains [6]. The TIR domain of *A. thaliana* (NM_123051) TIR proteins were downloaded from the TAIR database (http://www.arabidopsis.org/; 7), and used for tBLASTn search for similar sequences in the *C. cajan* genome sequence (BioProject ID: PRJNA68667; ACCESSION: AFSP00000000.1). The hits obtained were curated for full-length gene prediction, and the full length of the TIR domain and the genes with TIR domains were predicted using Fgenesh tool of softberry server (https://www.softberry.com/).

Phylogenetic classification and Motif identification

Multiple sequence alignment of the predicted TIR genes of *C. cajan* was done using ClustalW [8]. Alignment file with extension. aln was further used for phylogenetic tree construction using MEGA5.0. To elucidate best topology of tree bootstrap test (1000 replicates) was done with MEGA5.0 [9]. Identification of motifs within identified TIR containing proteins from *C. cajan* was done using MEME (http://meme.nbcr.net/meme/; 10). All parameters were set manually for motif elucidation.

Cis-acting Elements and *in-silico* Expression Study of Full Length Genes

Cis-acting elements present in the 1,000 bp sequence located upstream of each full-length TIR domain containing gene were determined. BLAST flavor tBLASTn [11] was used for retrieval of the 1,000 bp upstream sequences. PLANTCARE (http://bioinformatics. psb. ugent. be/webtools/plantcare/html/; 12) and PLACE (http:// www. dna. affrc. go.jp/PLACE/; 13) were used for *cis*-acting elements analysis. Computation level expression study of the hypothetical TIR proteins encoded by the predicted TIR genes of *C. cajan* was carried out by using *Glycine max* gene expression atlas (http://soybase.org/soyseq/; 14) as a reference dataset.

Homology Modeling of TIR Domain and Assessment of the Predicted Model

A candidate TIR domain (AFSP01036179_5692_6123) protein from *C. cajan* was taken for homology modeling using discovery studio 3.5 [15]. PDBSum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html;16) and ProSA (https://prosa.services.came.sbg.ac.at/prosa.php;17) servers were used for structural assessment of the predicted model.

Plant Materials and PCR Product Amplification

Seeds of the 34 pigeonpea genotypes and four of its wild relatives differing for reaction to Fusarium wilt [Table-1] were germinated, and about 100 mg (fresh weight) of healthy leaf sample was collected from 21 days old seedlings of each genotype. DNA extraction was carried out using the DNA extraction kit (Quiagen, Germany). DNA quality was visualized by agarose (0.8%) gel electrophoresis and its quantity was determined using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). The TIR domain based primers were synthesized from Operon Technologies, California, USA. One pair of primers [Table-2] specific for the candidate TIR domain was used for PCR amplification with the thirty eight genotypes.

Table 1- The pigeonpea genotypes and some of its wild relatives used for PCR amplification: their characteristics features and reaction to Fusarium wilt pathogen *Fusarium udum*.

S. No.	Genotype	Fusarium wilt Reaction*	Characteristic Features
1	Bahar	S	Compact, yellow flower, purple pod
2	IPA-204	R	Semi-compact, tall, yellow flower, pod green with streaks
3	KPL-43	R	Compact, yellow flower, purple pod
4	BDN-2010	R	Semi-compact, reddish yellow flower with red base, purple pod
5	BDN-2029	R	Semi-compact, dark red flower, pod green with purplish streaks
6	IPA-8F	R	Semi-spreading, dark red flower, purple pod
7	BDN-2001-9	R	Spreading, yellow flower, pod green with streaks
8	IPA-234	R	Compact, yellow flower, pod green with streaks
9	ICP-9174	MR	Semi-spreading, yellow flower, pod green with streaks
10	BWR-23	R	Flower red, pod green with streaks
11	BSMR-846	R	Semi-compact, yellow flower with red streaks, pod green with streaks
12	IPA-9F	R	Compact, light yellow flower with light red base, pod green with streaks
13	IPA-16F	R	Spreading yellow flower, tall, pod green with streaks
14	BDN-2004-1	R	Semi-spreading, medium dwarf, purple flower, pod green with streaks, purple stem
15	BWR-133	R	Semi-spreading, yellow flower, pod green with streaks
16	MAL-31	S	Semi-spreading, light yellow flower, pod green with streaks, brown bold seed (14g/100 seeds)
17	NDA-1	MS	Compact, yellow flower with purple veins, green pod
18	ICP-9150	MR	Compact, purple stem, yellow flower, pod green with streaks
19	Amar	MR	Compact, yellow flower, purple pod with constricted locule
20	MAL-13	MR	Spreading, light yellow flower, pod green with streaks and constricted locule, resistant to sterility mosaic virus
21	MA-6	MR	Semi-spreading, yellow flower, purple pod, highly resistant to sterility mosaic virus
22	MAL-18	S	Spreading, yellow flower, purple pod, highly resistant to sterility mosaic virus
23	ICP-2376	MS	Semi-spreading , yellow flower, pod green with streaks
24	LRG-41	MR	Semi-spreading, purple pod
25	BSMR-301	MR	-
26	MAL-23	MS	Spreading, yellow flower, purple pod, susceptible to sterility mosaic virus
27	MA-3	MS	Semi- spreading, yellow flower, small green pod with streaks and constricted locules , small seeds (9g/ 100 seeds)
28	MAL-34	MS	Spreading, red flower, pod green with streaks
29	MA Deo-89	S	Compact, flower yellow with purple streaks
30	MA PTH-2	S	Compact, red flower and pod, large red seeds, highly resistant to sterility mosaic virus
31	JKM 7	S	Compact, flower yellow with purple streaks
32	ICP 11887	S	Spreading, yellow flower
33	ICP 8860	-	-
34	ICP 8863	S	Highly resistant to sterility mosaic virus
35	*C. scarabaeoides*	R	Trailing habit, small pods with small black seeds (< 3g/100 seeds), resistant to pod fly and pod borer
36	*C. cajanifolius*	MS	Semi-spreading, yellow flower, pod green, small black seeds
37	*C. volubilis*	R	-
38	*Rhynchosia species*	R	-

*Wilt reaction based on two years (2009-10 and 2010-11) of screening with *Fusarium udum*. R, resistant; MR, moderately resistant; S, susceptible.

Table 2- Primer sequence for TIR domain amplification of the candidate gene (AGCT01041475.1_9515_9937).

Primer	Sequence	Length	Tm (°C)	GC (%)	Product size (bp)
TIR1F	TGCATCAGAAACCAAAGCAG	20	59.99	45	606
TIR1R	TTCTTCATGTTGTGGCCTGA	20	60.24	45	

The optimized 25µL PCR reaction mixture contained 200µM dNTPs, 1.0 × reaction buffer, 1.5mM $MgCl_2$, 1.25µM of each primer, 1.25U of *Taq* DNA polymerase, and 25ng of genomic DNA. All the PCR ingredients were obtained from Fermentas, USA. The amplification was done in a thermal cycler (Touchgene Gradient, Techne, UK) programmed for 40 cycles with an initial denaturation of 5 min at 94°C, followed by 39 cycles of 30 s denaturation at 94°C, 30 s annealing at 55°C, 1 min extension at 72°C and a final extension step of 7 min at 72°C. The PCR products were separated on 2.5% agarose gels in 1.0 × TBE buffer at 70 V using a horizontal gel electrophoresis system (Banglore Genei, India) for 3 h, stained with

ethidium bromide and photographed by a gel documentation system (Alpha Imager 2200, Alpha Innotech Corporation, San Leandro, USA).

Sequencing and Verification

PCR product was used for sequencing and further verification was done using bioinformatics tool and databases. BLAST tool was use for similarity search against NCBI nr database.

Results

In-silico Analysis of TIR Proteins from *C. cajan*

Based on homology search of the *C. cajan* genome using Arabidopsis TIR domain as query sequence, a total of ~119 hits were obtained using genome assembled by Singh Nagendra K. et al., [18]. Curation of these hits permitted tentative identification of 103 TIR domain containing genes. Functional domain analysis using Interproscan revealed that ~ 60 of these genes had TIR and NB-ARC domains, 21 genes had TIR, NB-ARC and LRR domains, while 22 genes had only TIR domain [Table-3] and [Table-4].

Phylogenetic Classification

The predicted 103 TIR domain containing *C. cajan* genes were subjected to phylogenetic classification using multiple sequence alignment and tree construction by MEGA 5.1. Three major groups were observed: the first group had 61 members and was divided into two subgroups; the second group had 37 members and was also divided into two subgroups, while the third group was rather small and homogeneous with only five members [Fig-1]. Further analyses were restricted to the 21 TNL genes (TIR genes having all the three, *viz.*, TIR, NBS and LRR domains); the remaining 82 genes were not investigated further. Phylogenetic classification of these 21 TNL genes showed that they could be divided into two main groups of 14 and seven genes; the larger group was further divided into two subgroups of 9 and 5 genes [Fig-2].

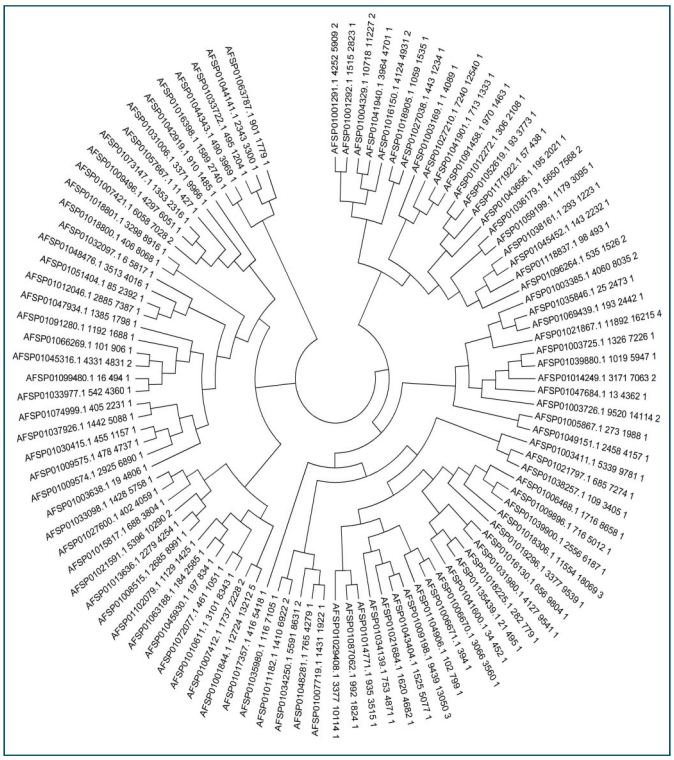

Fig. 1- Phylogenetic classification of TIR proteins from *C. cajan*.

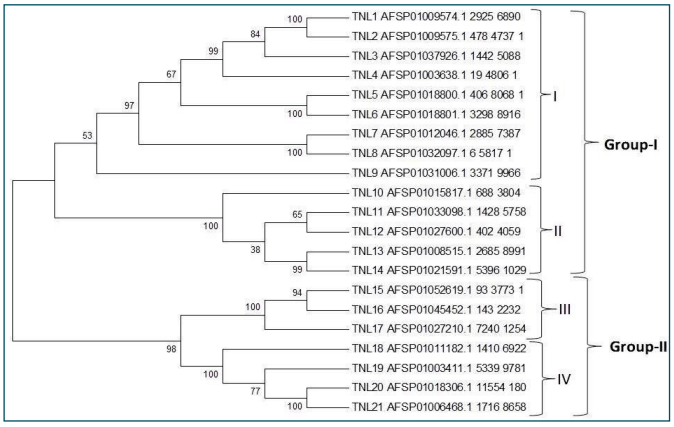

Fig. 2- Phylogenetic tree of proteins encoded by the 21 TNL genes (genes having TIR-NBS-LRR domains) from *C. cajan*.

Motif Identification

Motif analysis of the 21 TNL genes was carried out with MEME tool. All the 21 genes had at least one TIR, one NBS and one LRR domain. Further, four of the genes *CcTNL*9, *CcTNL*14, *CcTNL*15 and *CcTNL*19 had two copies of the NBS domain; in the case of two of these genes (*CcTNL*14 and *CcTNL*19), the additional NBS domain was located next to the TIR domain, while in the case of genes *CcTNL*9 and *CcTNL*15 they were situated close to the C-terminus beyond the LRR domain repeats [Fig-3],[Fig-4]. These motif sequences were further validated with Motif Scan server. Motif Scan results revealed that identified motifs are having major similarity with respective domains.

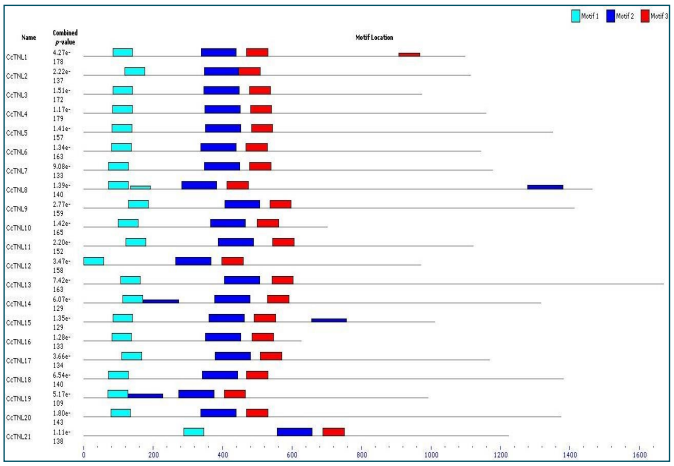

Fig. 3- Identified Motif Representation

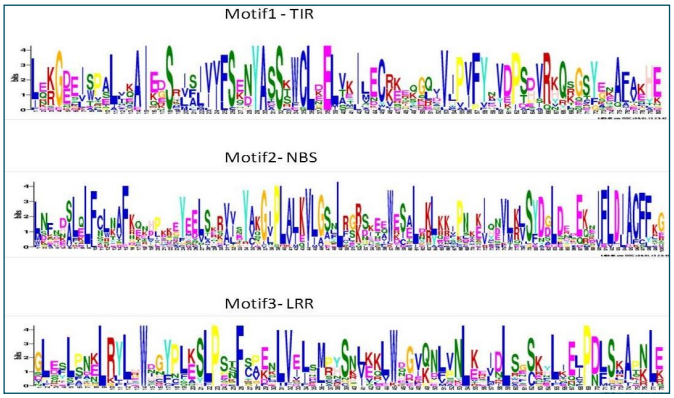

Fig. 4- Motif details with their Sequence Logo representation.

Complete gene organization

The complete gene structures for the 21 TNL genes were successfully determined using Fgenesh. The details, including number of exons and complete CDS length of each gene are shown in [Fig-5] and summarized in [Table-3] and [Table-4]. The number of exons ranged from merely 3 (*CcTNL*16) to 16 (*CcTNL*05), while the CDS ranged from 1887bp (*CcTNL*16) to 5016bp (*CcTNL*13). The genomic gene size showed considerable variation from 2090bp (*CcTNL*16) to 7663bp (*CcTNL*05). Similarly, the size of protein product varied from 629aa to 1671aa [Table-4].

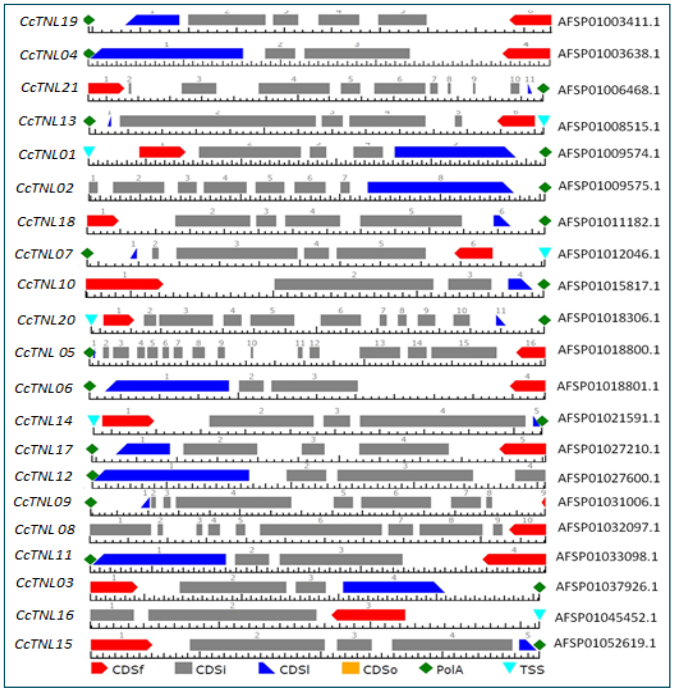

Fig. 5- Complete gene organization for the 21 TNL genes.

Table 3- Gene specific Primary Accession Numbers with initiation codon and termination codon positions, exons and CDS length details for the 21 full-length TNL genes.

Gene Name	Primary Accession Number	Start Position	End Position	Number of Exons	CDS Length
CcTNL21	AFSP01006468.1	1716	8658	11	3678bp
CcTNL13	AFSP01008515.1	2685	8991	6	5016bp
CcTNL01	AFSP01009574.1	2925	6890	5	3303bp
CcTNL18	AFSP01011182.1	1410	6922	6	4362bp
CcTNL07	AFSP01012046.1	2885	7387	6	3540bp
CcTNL20	AFSP01018306.1	11554	18069	11	4131bp
CcTNL14	AFSP01021591.1	5396	10290	5	3960bp
CcTNL03	AFSP01037926.1	1442	5088	4	2931bp
CcTNL16	AFSP01045452.1	143	2232	3	1887bp
CcTNL09	AFSP01031006.1	3371	9966	9	4248bp
CcTNL11	AFSP01033098.1	1428	5758	4	3375bp
CcTNL10	AFSP01015817.1	688	3804	4	2115bp
CcTNL08	AFSP01032097.1	6	5817	10	4398bp
CcTNL17	AFSP01027210.1	7240	12540	5	3513bp
CcTNL04	AFSP01003638.1	19	4806	4	3483bp
CcTNL05	AFSP01018800.1	406	8068	16	4062bp
CcTNL19	AFSP01003411.1	1	5339	6	2985bp
CcTNL15	AFSP01052619.1	93	3773	5	3042bp
CcTNL12	AFSP01027600.1	402	4059	4	2919bp
CcTNL02	AFSP01009575.1	478	4737	8	3351bp
CcTNL06	AFSP01018801.1	3298	8916	4	3438bp

Table 4- Gene abbreviation with genomic and protein length and available upstream region.

S. No.	Gene Name	Gene Detail (Primary Accession No)	Upstream Sequence (bp)	Protein Length (aa)	Genomic Length (bp)
1	*CcTNL21*	AFSP01006468.1	1000	1225	6943
2	*CcTNL13*	AFSP01008515.1	1000	1671	6307
3	*CcTNL01*	AFSP01009574.1	1000	1100	3966
4	*CcTNL18*	AFSP01011182.1	1000	1384	6181
5	*CcTNL07*	AFSP01012046.1	1000	1179	4503
6	*CcTNL20*	AFSP01018306.1	1000	1376	6516
7	*CcTNL14*	AFSP01021591.1	1000	1319	4895
8	*CcTNL03*	AFSP01037926.1	1000	976	3647
9	*CcTNL16*	AFSP01045452.1	1000	629	~2090
10	*CcTNL09*	AFSP01031006.1	961	1415	6596
11	*CcTNL11*	AFSP01033098.1	864	1124	4331
12	*CcTNL10*	AFSP01015817.1	687	704	3117
13	*CcTNL08*	AFSP01032097.1	578	1466	~5817
14	*CcTNL17*	AFSP01027210.1	488	1170	5301
15	*CcTNL04*	AFSP01003638.1	293	1160	4788
16	*CcTNL05*	AFSP01018800.1	236	1353	7663
17	*CcTNL19*	AFSP01003411.1	258	994	4443
18	*CcTNL15*	AFSP01052619.1	92	1013	3681
19	*CcTNL12*	AFSP01027600.1	-	972 (partial)	~4737
20	*CcTNL02*	AFSP01009575.1	-	1116 (partial)	~3694
21	*CcTNL06*	AFSP01018801.1	-	1145	5619

In-silico Analysis of Putative Promoter Region of TNL Genes

A total of 1000bp upstream sequence of each of the 21 TNL genes was retrieved, depending on the availability of the genome sequence in the database. In the case of 9 genes, 1,000 bp upstream sequences were available, no upstream sequence was found for three genes (*CcTNL*12, *CcTNL*02 and *CcTNL*06), and for the remaining nine genes less than 1,000 bp upstream sequence could be retrieved. Out of the three genes for which upstream sequences could not be retrieved, in two of the genes (*CcTNL*12 and *CcTNL*02) the initiation codon was absent, while the third gene (*CcTNL*06) had the initiation codon. Three major groups of elements were identified in the upstream regions of the 21 *CcTNL* genes, *viz.,* (1) hormone response elements (2) seed-specific elements and (3) biotic and abiotic stress response elements. Gibberellin response elements (TATC-box and GARE motif) were found in *CcTNL*11 and *CcTNL*20 upstream regions. Methyl jasmonate responsive elements (CGTCA and TGAGG motifs) were found in the upstream regions of *CcTNL*01, *CcTNL*20, *CcTNL*16, *CcTNL*09, *CcTNL*08 and *CcTNL*07. TCA element is involved in salicylic acid responsiveness; it was observed in putative promoter regions of *CcTNL*13 and *CcTNL*14. The ABRE element involved in abscisic acid responsiveness was detected in *CcTNL*01 and *CcTNL*20 upstream regions. Ethylene response element ERE was found in the upstream region of only one gene *CcTNL*16 [Table-5]. SKN1, GCN4 and RY elements are involved in endosperm-specific gene expression; these elements were found in the upstream regions of eight of the 21 genes, *viz., CcTNL*13, *CcTNL*01, *CcTNL*07, *CcTNL*20, *CcTNL*14, *CcTNL*03, *CcTNL*16 and *CcTNL*09 genes [Table-6]. Biotic and abiotic stress response element TC-rich motif was present in the putative promoter regions of eight of the *CcTNL* genes, viz., *CcTNL*21, *CcTNL*18, *CcTNL*20, *CcTNL*14, *CcTNL*09, *CcTNL*04, *CcTNL*05 and *CcTNL*19. The MBS element was located in the upstream sequences of genes *CcTNL*21, *CcTNL*20, *CcT-*

*NL*11, *CcTNL*10 and *CcTNL*17. Box-W1 was detected in the up-stream region of gene *CcTNL*01 and HSE element was found in upstream of *CcTNL*18, *CcTNL*07, *CcTNL*20, *CcTNL*16, *CcTNL*09, *CcTNL*10 and *CcTNL*17 genes [Table-7].

Table 5- Hormone responsive elements availability in 21 genes.

Elements	Function	CcTNL21	CcTNL13	CcTNL01	CcTNL18	CcTNL07	CcTNL20	CcTNL14	CcTNL03	CcTNL16	CcTNL09	CcTNL11	CcTNL10	CcTNL08	CcTNL17	CcTNL04	CcTNL05	CcTNL19	CcTNL15	CcTNL12	CcTNL02	CcTNL06
TATC-box	Cis-acting element involved in gibberellin-responsiveness	Y					Y															
TCA-element	Cis-acting element involved in salicylic acid responsiveness		Y					Y														
ABRE	Cis-acting element involved in the abscisic acid responsiveness			Y			Y						Y				Y					
CGTCA-motif	Cis-acting regulatory element involved in the MeJA-responsiveness			Y			Y			Y	Y			Y								
ERE	Ethylene-responsive element									Y												
GARE-motif	Gibberellin-responsive element											Y										
TGACG-motif	Cis-acting regulatory element involved in the MeJA-responsiveness					Y				Y								Y				
TGA-element	Auxin-responsive element												Y		Y							

Table 6- Seed specific elements availability in 21 genes.

Elements	Function	CcTNL21	CcTNL13	CcTNL01	CcTNL18	CcTNL07	CcTNL20	CcTNL14	CcTNL03	CcTNL16	CcTNL09	CcTNL11	CcTNL10	CcTNL08	CcTNL17	CcTNL04	CcTNL05	CcTNL19	CcTNL15	CcTNL12	CcTNL02	CcTNL06
Skn-1_motif	cis-acting regulatory element required for endosperm expression		Y	Y		Y	Y (3)	Y		Y (2)	Y (2)											
GCN4_motif	cis-regulatory element involved in endosperm expression							Y														
RY-element	cis-acting regulatory element required for endosperm expression		Y					Y	Y													

Table 7- Biotic and abiotic specific elements availability in 21 genes.

Elements	Function	CcTNL21	CcTNL13	CcTNL01	CcTNL18	CcTNL07	CcTNL20	CcTNL14	CcTNL03	CcTNL16	CcTNL09	CcTNL11	CcTNL10	CcTNL08	CcTNL17	CcTNL04	CcTNL05	CcTNL19	CcTNL15	CcTNL12	CcTNL02	CcTNL06
TC-rich repeats	Cis-acting element involved in defense and stress responsiveness	Y			Y		Y (3)	Y			Y					Y	Y	Y				
MBS	MYB binding site involved in drought-inducibility	Y					Y					Y (2)	Y		Y (2)							
Box-W1	fungal elicitor responsive element			Y																		
HSE	cis-acting element involved in heat stress responsiveness				Y	Y	Y			Y	Y		Y		Y							
WUN-motif	wound-responsive element														Y							
AT-rich sequence	element for maximal elicitor-mediated activation											Y(2)										
Y = available																						

Table 8- Plant ontology for the 14 tissues of *Glycine max*.

Tissue	DAF	Ontology Term	Ontology Identifier
Young Leaf	NA	0.4 leaflets unfurled	Soy:000252
Flower	NA	F0.4 Open flower	Soy:001277
One cm pod	7 DAF	F0.7 Small size pod	Soy:001280
Pod-shell	10-13 DAF	F0.8 Pod medium size	Soy:001281
Pod-shell	14-17 DAF	F0.9 Full pod size	Soy:001282
Seed	10-13 DAF	S1.06 Cotyledon Stage	Soy:001290
Seed	14-17 DAF	S1.06 Cotyledon Stage	Soy:001290
Seed	21 DAF	S1.07 Early maturity stage 1	Soy:001291
Seed	25 DAF	S1.07 Early maturity stage 1	Soy:001291
Seed	28 DAF	S1.07 Early maturity stage 1	Soy:001291
Seed	35 DAF	S1.08 Early maturity stage 2	Soy:001292
Seed	42 DAF	S1.09 Mid seed maturity	Soy:001293
Root	NA	Root structures	Soy:001183
Nodule	NA	Bacterial root nodule	Soy:001301

In- Silico Expression Analysis of the 21 Full-length Genes

In view of the close phylogenetic relationship between *C. cajan* and *G. max,* the expression profiles of *G. max* TIR associated genes were used for in silico expression analysis of the 21 *CcTNL* genes. The sequences of the 21 *CcTNL* genes were compared with those of the *G. max* TIR associated genes for homology and the phylogenetic tree was constructed using MEGA 5.0. The *G. max* gene(s) with the highest similarity score and falling in the same clade as the *CcTNL* genes were identified and their expression profiles were retrieved from the transcriptome datasets available in Soybase. The soybean RNA Seq-Atlas provides high-resolution gene expression in a diverse set of fourteen tissues [Table-8].

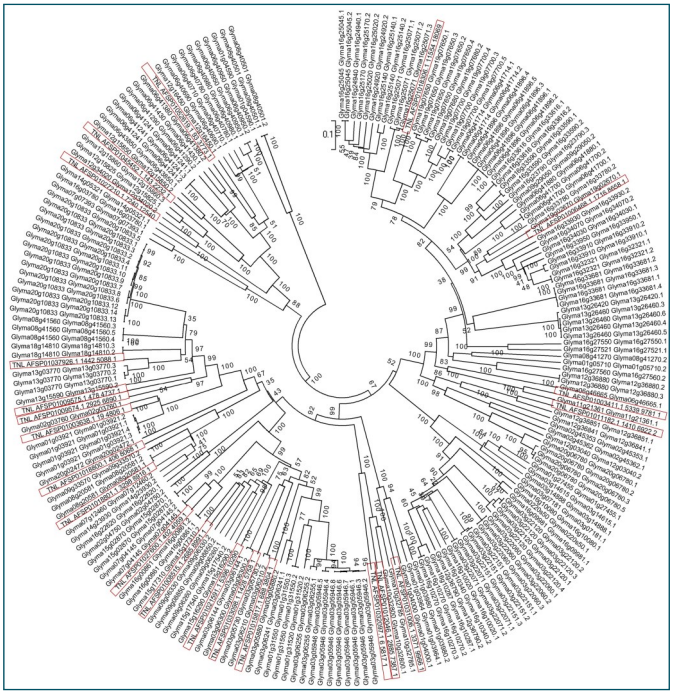

Fig. 6- Comparative phylogenetic analysis with *Glycine max* and *Cajanus Cajan*

Mining of these data suggested three clades of tissues, *viz.,* aerial (leaf and flower) underground (root and root nodule) and seed, in

which the individual *CcTNL* genes exhibited markedly different expression patterns. The predicted expression profiles of the *CcTNL* genes on the basis of homologous *G. max* TIR associated genes are depicted in [Fig-7] and summarized in [Table-9].

This *in silico* investigation was able to predict the expression pattern of 10 of the 21 *CcTNL* genes, while the expression profile of the remaining 11 genes could not be predicted on the basis of the soybean transcriptome dataset [Fig-7].

Table 9- Comparative table of *Cajanus cajan* and *Glycine max* with expression signatures availability with 14 tissues.

Gene Name in Cc	Gene Name in Gm	Young leaf	flower	one cm pod	pod shell 10DAF	pod shell 14DAF	seed 10DAF	seed 14DAF	seed 21DAF	seed 25DAF	seed 28DAF	seed 35DAF	seed 42DAF	root	nodule
CcTNL10_AFSP01015817.1_688_3804	Glyma03g06241_Glyma03g06241.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL11_AFSP01033098.1_1428_5758	Glyma03g06144_Glyma03g06144.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL12_AFSP01027600.1_402_4059	Glyma07g04145_Glyma07g04145.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL13_AFSP01008515.1_2685_8991	Glyma15g17310_Glyma15g17310.2	0	1	0	0	0	0	0	0	0	1	8	2	12	7
CcTNL14_AFSP01021591.1_5396_10290	Glyma15g17540_Glyma15g17540.2	0	0	0	0	0	1	0	0	0	0	0	0	0	0
CcTNL15_AFSP01052619.1_93_3773	Glyma12g16450_Glyma12g16450.1	100	61	84	95	95	37	29	14	34	24	44	20	30	30
CcTNL16_AFSP01045452.1_143_2232	Glyma12g15860_Glyma12g15860.3	17	71	11	21	14	7	0	1	2	3	2	4	30	14
CcTNL04_AFSP01003638.1_19_4806	Glyma02g03760_Glyma02g03760.1 Glyma01g03921_Glyma01g03921.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL01_AFSP01009574.1_2925_6890	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL02_AFSP01009575.1_478_4737	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL03_AFSP01037926.1_1442_5088	Glyma13g03770_Glyma13g03770.1	0	13	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL05_AFSP01018800.1_406_8068	Glyma09g33570_Glyma09g33570.2	0	0	0	0	0	0	0	0	0	0	0	0	1	0
CcTNL06_AFSP01018801.1_3298_8916	Glyma14g23930_Glyma14g23930.2	6	2	2	8	16	1	2	0	0	0	0	0	14	7
CcTNL09_AFSP01031006.1_3371_9966	Glyma01g04000_Glyma01g04000.1	26	103	19	22	16	2	2	3	9	2	11	3	59	0
CcTNL07_AFSP01012046.1_2885_7387	Glyma10g32800_Glyma10g32800.1	3	13	13	19	20	2	0	1	1	2	1	0	1	2
CcTNL08_AFSP01032097.1_6_5817	-	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL19_AFSP01003411.1_5339_9781	Glyma11g21361_Glyma11g21361.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL20_AFSP01018306.1_11554_18069	Glyma19g07650_Glyma19g07650.1	1	22	1	2	4	1	0	0	1	2	5	0	6	2
CcTNL21_AFSP01006468.1_1716_8658	Glyma19g02670_Glyma19g02670.1	13	26	16	26	10	0	1	0	6	0	4	0	1	2
CcTNL18_AFSP01011182.1_1410_6922	Glyma12g36851_Glyma12g36851.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CcTNL17_AFSP01027210.1_7240_12540	Glyma12g34020_Glyma12g34020.2	4	3	6	3	0	0	2	0	0	0	0	0	24	2

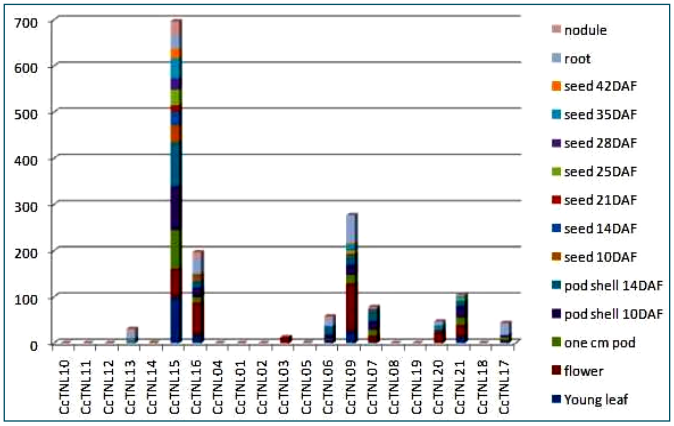

Fig. 7- *In silico* expression analysis details of the 21 *CcTNL* genes on the basis of the expression profiles of their homologous G. max TIR associated genes.

Amplification and Sequencing of TIR domain

To verify the TIR domain consistency, genomic DNAs from 38 cultivars of *C. cajan* differing in their response to Fusarium wilt (FW) and four of its wild relatives were used for amplification with TIR domain-specific primers. A single was obtained in all the cases, except for the wild species *C. cajanifolius*, which did not yield any. The TIR domain amplification products [Fig-8] from only three cultivars, namely, BSMR-846 (resistant to FW), MAL-13 (moderately resistant to FW) and Bahar (susceptible to FW) were used for sequencing. In case of Bahar, more intense band was obtained. After sequencing of the PCR products, the annotated sequences were successfully deposited in NCBI database with Accession Nos. JX667781.1 (BSMR-846), JX667782.1 (MAL-13) and JX667783.1 (Bahar).

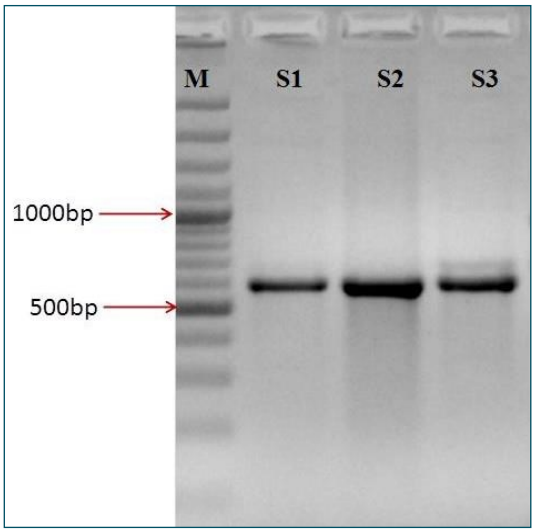

Fig. 8- Primer TIR 1 amplified a PCR product of approx. 606 bp in the pigeonpea genotypes BSMR-846 (S1; FW resistant) and MAL-13 (S3; FW moderately resistant); however it amplified a more intense band in the pigeonpea genotype, Bahar (S2; FW susceptible).

Homology Modeling of TIR Domain and Refinement

Since there was no structure available for *C. cajan* TIR domain in PDB structure database, an approach homology modeling was done for functional analysis of this domain. Crystal structure of TIR Domain from *A. thaliana* (3JRN.pdb) was used as a template for 3D

prediction of *C. cajan* TIR Domain (AFSP01036179_5692_6123). Sequence identity between template and target was 46% with 64% positive similarities. The best predicted model was used for structural verification; it was found that 96.5% of the residues were present in favoured region, 3.5% of the residues (ASN21, THR92, SER112, GLU124 and ASP129) were placed in the allowed region, and no residue was found in the disallowed region [Fig-9].

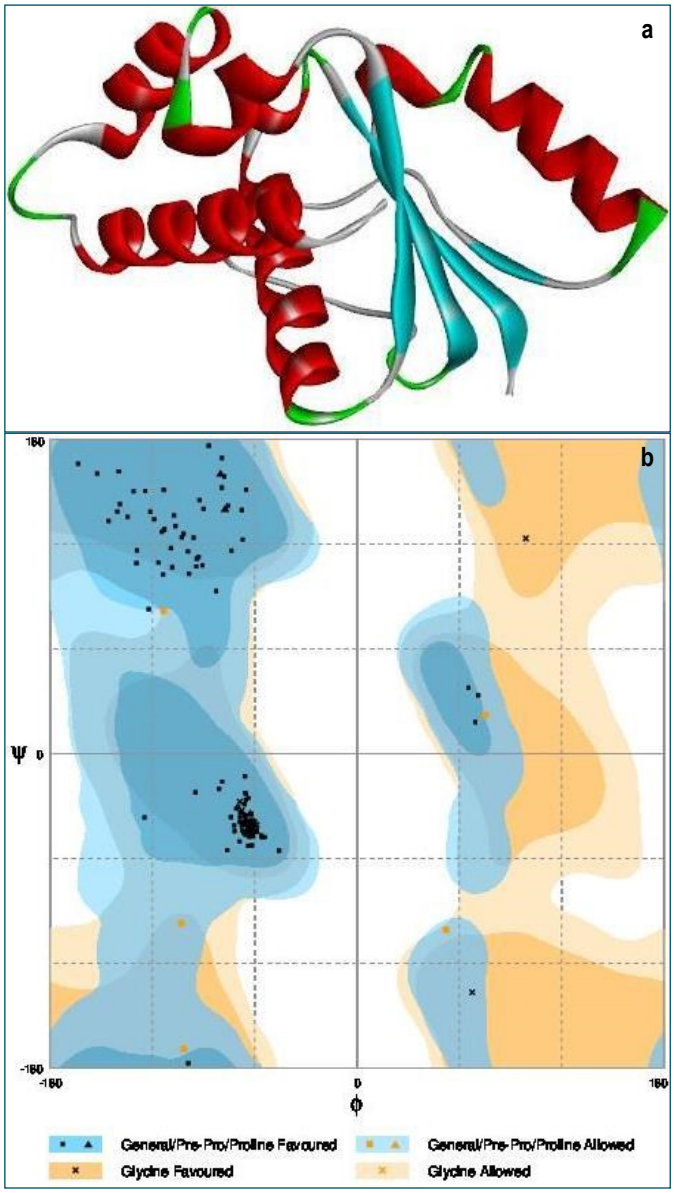

Fig. 9- (a) Three dimensional structure of TIR domain form *C. cajan.* **(b)** Stereo-chemical quality statistics using RAMPAGE server.

This best reliable model of the TIR domain from *C. cajan* was successfully deposited in the PMDB database with PMDBID PM0078097. The overall model quality score for target was -5.55 and for template -7.4 z-score. ERRAT based quality factor was 75%. Secondary elements statistics showed that 3D structure contains 46% alpha-helices, 20% beta-sheets, 32% coils and 22% turns. Hydrogen bond statistics indicated that mean H bond distance was 2.2 (sd = 0.3) against the expected value of 2.2 (sd = 0.4). Mean H bond energy was observed as -1.6 (sd = 1.0) against the expected value of -2.0 (sd = 0.8). It was found that 77% of the residues had participated in H bond formation, which was slightly more than the expected 75% of the residues. For prediction of resolution, RESPROX server was used, and it was found as 2.279Å, which shows the medium resolution structure. Based on prosa server, the target score was -5.55 against the template score of -7.4. All the above structural investigations suggest that the predicted model is reliable with good quality score [Fig-10].

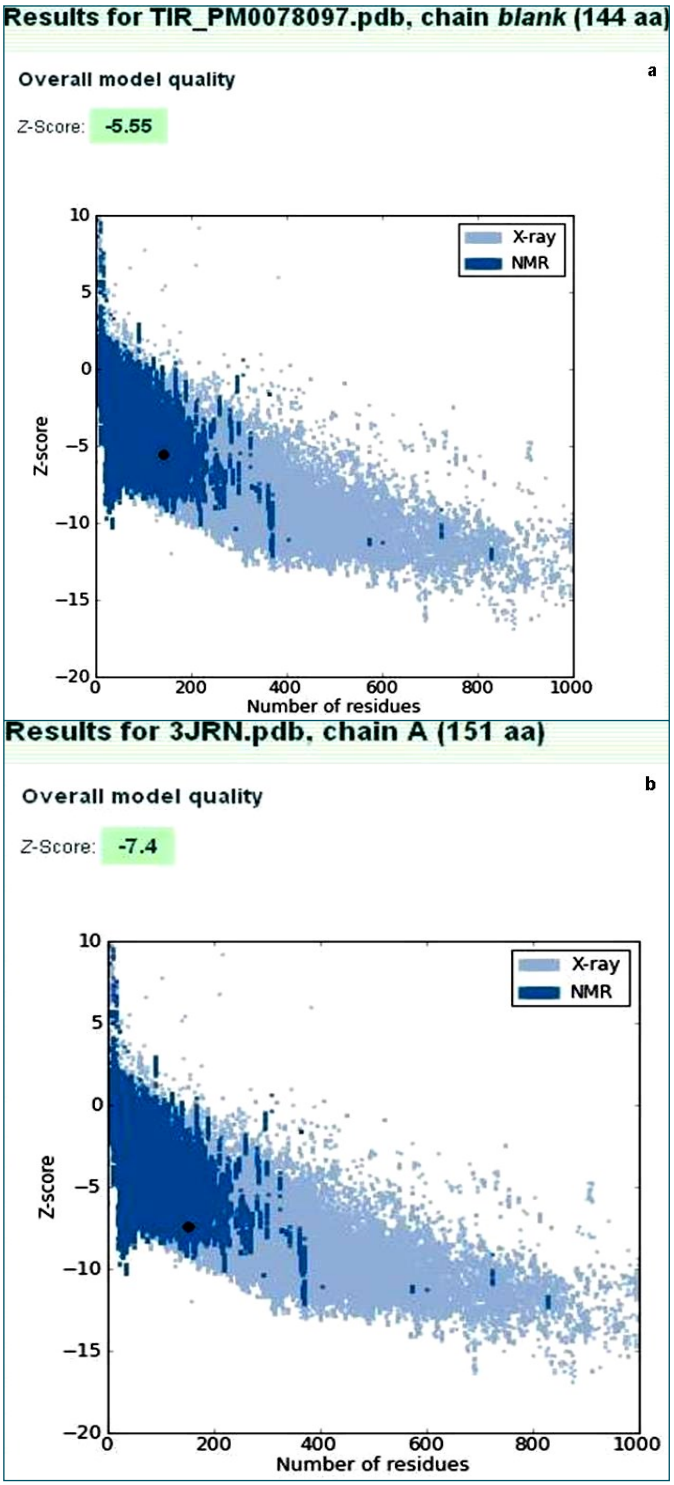

Fig. 10- ProSA structure quality analysis for target **(a)** and template **(b)** structure.

Discussion

Based on the available genome sequence of *C. cajan*

(AFSP01000000), near about 119 hits were found similar with TIR domain associated family members, of which 103 genes had TIR domains in their genic region (http://www.insilicogenomics.in/r/cc.html). After complete analysis, 21 genes were found to have TIR, NBS and LRR domains; these 21 genes were subjected to *in silico* characterization. The genes having TNL (TIR, NBS and LRR domain) were classified on the basis of maximum similarity algorithm (UPGMA method; MEGA software) into two groups, both the groups having two subgroups each. Based on motif identification 3 major motifs were observed in which motif I belongs to TIR domain (IPR000157), motif II belongs to NB-ARC (IPR002182) and P-loop containing nucleoside triphosphate hydrolase (IPR027417), while motif III belongs to LRR domain confirmed by signature recognition search using Interproscan server. Complete gene organization revealed that maximum gene size length was 5,016bp with 1671aa. Further, hormone responsive elements, seed-specific elements, and biotic and abiotic response elements were successfully identified in the upstream regions of the 21 genes. Based on *cis*-acting elements study, it has been found that Box-W1 fungal elicitor responsive element was only observed in only *CcTNL*01 gene (AFSP01009574.1: 2925-6890) upstream region. In case for hormone responsive elements specific gene, abscisic acid, auxin, ethylene, gibberellin, MeJA and salicylic acid response elements were found in different genes, and seed-specific elements (Skn-1_motif, GCN4_motif and RY-element) were observed in different genes, while in case of biotic and abiotic stress TC-rich repeats, MBS, WUN-motif and AT-rich sequences were successfully observed. *in silico* expression study based on *G. max* expression datasets, only 5 genes showed comparative match with *C. cajan* genes (AFSP01052619.1: 93-3773, AFSP01045452.1: 143_2232, AFSP01031006.1: 3371-9966, AFSP01012046.1: 2885-7387 and AFSP01018306.1: 11554-18069).

A candidate TIR domain was taken for amplification in *C. cajan* genome and successfully amplified 606bp product with genomic DNA of pigeonpea genotypes BSMR-846 (FW resistant), MAL-13 (moderately resistant to FW) and Bahar (FW susceptible) and complete annotated sequences were deposited in NCBI database with Accession Nos. JX667781.1, JX667782.1 and JX667783.1 and corresponding protein ID were AFV93474.1, AFV93475.1 and AFV93476.1, respectively. After getting protein sequences, structural classification was done to elucidate the functional role of amplified TIR domain. Since no structure was available for *C. cajan* TIR domain in structure database, homology modeling was done. Crystal Structure of *A. thaliana* TIR domain (3JRN.pdb) was used as a template and the best model of *C. cajan* TIR domain was successfully deposited in the Protein Model Database with PMDBID PM0078097. Structural classification revealed that this domain having classification lineage 3.80.10.10, having α-β class with Alpha-Beta barrel architecture and TIM Barrel, Rossmann fold type topology. Over all sequential and structural studies implicate that TIR domain has major role in self-association, signaling, and auto-regulation with respect to plant immunity and defence.

Acknowledgements

V.K. Singh is thankful to the Center of Bioinformatics, School of Biotechnology, Banaras Hindu University, Varanasi, and Department of Biotechnology, Government of India, New Delhi, India, for *in silico* facilities and financial support. He is also thankful to Ms. Sakshi Singh, M.Sc. Bioinformatics, Allahabad University for data management.

Conflicts of Interest: None declared.

References

[1] Burch-Smith T.M., Schiff M., Caplan J.L., Tsao J., Czymmek K. and Dinesh-Kumar S.P (2007) *PLoS Biol.*,5(3), e68.

[2] Swiderski M., Birker D. and Jones J. (2009) *Mol. Plant Microbe. Interact.*, 22, 157-165.

[3] Bernoux M., Ve T., Williams S., Warren C., Hatters D., Valkov E., Zhang X., Ellis J.G., Kobe B. and Dodds P.N. (2011) *Cell Host Microbe.*, 9(3), 200-11.

[4] Whitham S., Dinesh-Kumar S.P., Choi D., Hehl R., Corr C. and Baker B. (1994) *Cell,* 78, 1101-1115.

[5] Chan S.L., Mukasa T., Santelli E., Low L.Y. and Pascual J. (2010) *Protein Sci*., 19, 155-161.

[6] Jebanathirajah J.A., Peri S. and Pandey A. (2002) *Trends Plant Sci.*, 7(9), 388-391.

[7] Lamesch P., Dreher K., Swarbreck D., Sasidharan R., Reiser L. and Huala E. (2010) *Current Protocols in Bioinformatics*, 1-11.

[8] Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) *Bioinformatics*. 23(21), 2947-2948.

[9] Kumar S., Tamura K. and Nei M. (1994) *Comput. Appl. Biosci.,* 10(2), 189-191.

[10] Bailey T.L., Williams N., Misleh C. and Li W.W. (2006) *Nucleic Acids Res.,* 34369-34373.

[11] Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J. (1990) *J. Mol. Biol.,* 215, 403-410.

[12] Rombauts S., Déhais P., Van Montagu M. and Rouzé P. (1999) *Nucleic Acids Res.,* 27, 295-296.

[13] Higo K., Ugawa Y., Iwamoto M. and Korenaga T. (1999) *Nucleic Acids Res.* 27(1), 297-300.

[14] Severin A.J., Woody J.L., Bolon Y.T., Joseph B., Diers B.W., Farmer A.D., Muehlbauer G.J., Nelson R.T., Grant D., Specht J.E., Graham M.A., Cannon S.B., May G.D., Vance C.P. and Shoemaker R.C. (2010) *BMC Plant Biol.,* 5, 160.

[15] Shahi S.K., Singh V.K. and Kumar A. (2013) *PLoS ONE*, 8, 68234.

[16] Laskowski R.A. (2008) *Nucleic Acids Res.,* 37, 355-359.

[17] Wiederstein M. and Sippl M.J. (2007) *Nucleic Acids Res.,* 35, 407-410.

[18] Singh N.K., Gupta D.K., Jayaswal P.K., Mahato A.K., Dutta S., Singh S. and Sharma T.R. (2012) *Journal of Plant Biochemistry and Biotechnology*, 21(1), 98-112.