Genome-wide identification and tissue-specific expression analysis of nucleotide binding site-leucine rich repeat gene family in *Cicer arietinum* (kabuli chickpea)

Ranu Sharma*,†, Vimal Rawat*,†, C.G. Suresha

*Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, Maharashtra, India
†Institute of Plant and Microbial Biology and Zurich-Basel Plant Science Center, University of Zurich, Zollikersstrasse, 107, CH 8008, Zurich, Switzerland

**A R T I C L E   I N F O**

**Keywords:**
*Cicer arietinum*
NBS-LRR
Gene annotation
Synteny
DNA-motifs
Gene expression

**A B S T R A C T**

The nucleotide binding site-leucine rich repeat (NBS-LRR) proteins play an important role in the defense mechanisms against pathogens. Using bioinformatics approach, we identified and annotated 104 NBS-LRR genes in chickpea. Phylogenetic analysis points to their diversification into two families namely TIR-NBS-LRR and non-TIR-NBS-LRR. Gene architecture revealed intron gain/loss events in this resistance gene family during their independent evolution into two families. Comparative genomics analysis elucidated its evolutionary relationship with other fabaceae species. Around 50% NBS-LRRs reside in macro-syntenic blocks underlining positional conservation along with sequence conservation of NBS-LRR genes in chickpea. Transcriptome sequencing data provided evidence for their transcription and tissue-specific expression. Four cis-regulatory elements namely WBOX, DRE, CBF, and GCC boxes, that commonly occur in resistance genes, were present in the promoter regions of these genes. Further, the findings will provide a strong background to use candidate disease resistance NBS-encoding genes and identify their specific roles in chickpea.

1. Introduction

Plants have evolved a multi-layered innate immune system to counter an enormous range of external adverse changes. The disease resistance genes (R-genes) play a critical role in plant defense mechanisms and respond to attack by several pathogens and pests, including viruses, bacteria, fungi, nematodes, and insects. The signaling component required during a defense response is decided by the R-gene structure [1].

One of the major classes of proteins encoded by R-gene family possesses the nucleotide binding site-leucine rich repeat (NBS-LRR) domains. The NBS domain has several conserved motifs that bind and hydrolyze ATP or GTP [2]. The LRR regions are involved in protein-protein interactions and thus play role in molecular recognition and signaling present, is known as non-TIR-NBS-LRR (non-TNL) or sometimes as CC-NBS-LRR (CNL) [7,8]. Moreover, the sequences of conserved motifs, especially those within the NBS domain, have been used extensively to identify novel disease resistance genes in the model and crop plants [9–11].

NBS-LRR resistance genes have been identified in gymnosperms to angiosperms [12]. Genome sequencing of the model plants has aided genome-level investigation of this gene family in monocot and dicot plant species such as *Oryza sativa* [13,14], *Malus domestica* [15,16], *Arabidopsis thaliana* [17,18], *Medicago truncatula* [19], *Zea mays* [20], *Carica papaya* [21], *Cucumis sativus* [22], *Brassica rapa* [23], *Populus trichocarpa* [24], *Vitis vinifera* [25], *Solanum tuberosum* [26], *Linum usitatissimum* L. [27], *Gossypium raimondii* [28], *Arachis duranensis*, *Arachis ipaensis* [29], *Actinidia chinensis* [30], and many more [31]. Previous studies have shown that NBS-LRR resistance genes constitute approximately 0.6 to 1.8% of the total genes encoded by plant genomes [26]. Moreover, it has been shown that the number of NBS-LRR genes is correlated with the total number of genes in the genome [22]. The number of NBS-LRR genes in different plant genomes varies substantially from < 100 to > 1.000 [12,32]. The largest number of resistance genes at present hold by *Nicotiana tabacum* (Eudicots) and

---

* Corresponding author at: Institute for Genetics, University of Cologne, Zülpicher Strasse 47a, 50674 Cologne, Germany.
E-mail address: rsharma1@uni-koeln.de (R. Sharma).
† These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gdata.2017.08.004
Received 10 May 2017; Received in revised form 5 August 2017; Accepted 9 August 2017
Available online 12 August 2017
2213-5960/ © 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Triticum aestivum (Monocots) [31]. Some other plant genomes such as M. truncatula, P. trichocarpa, and V. vinifera also encode for a large number of NBS-LRR resistance genes (333, 402, and 459 each). However, there are still numerous exceptions to this such as the presence of low copy number of NBS genes in C. papaya (54), B. rapa (92), C. sativus (57), and Z. mays (109).

According to the report by Food and Agriculture Organization (FAO) in 2008, chickpea is one of the oldest and second most widely grown crops in the world [33]. It is the primary source of human dietary nitrogen. However, many fungal diseases like Ascochyta blight (Ascochyta rabiei), Fusarium wilt (Fusarium oxysporum f. sp. ciceris), Botrytis gray mold (Botrytis cinerea), rust (Uromyces ciceris-arietini), and collar rot (Sclerotium rolfsii) lead to extensive crop damage affecting chickpea productivity [34] (www.icrisat.org/bt-pathology-fungal.htm). Majorly, the productivity of chickpea crop is drastically affected by two fungal diseases viz., Fusarium wilt and Ascochyta blight causing 100% loss in yield when conditions are favorable for infection.

Here, we report the results of an in silico study conducted to identify and characterize NBS-LRR resistance genes of chickpea genome. These findings will help to fish out candidate R-genes in chickpea and provide a blueprint for future efforts towards improvement of disease resistance in chickpea.

2. Results

2.1. Identification of NBS-LRR proteins

A blastp search using consensus sequences of TNL and non-TNL as query against the predicted chickpea proteomes resulted in the identification of 121 and 135 hits in CDC frontier (kabuli) and ICC4958 (desi) chickpea. Out of which, 104 and 119 sequences were true NBS-LRR proteins in the two varieties (based on the presence of NB-ARC domain) (Supplementary Fig. S1, Supplementary Table S1). The nomenclature used for naming these proteins/genes is according to the protein/gene identifiers given in the LIS database.

To confirm the above results, hidden Markov model profile search was carried out against the chickpea proteomes using HMMER (hmmssearch) and NB-ARC domain as query. A total of 172 and 201 hits were obtained, out of which 104 and 119 kabuli and desi chickpea protein sequences, identical dataset identified using blastp, were the true NBS-LRR proteins. These high quality NBS-LRR proteins were taken to build a chickpea-specific hidden Markov model to check for any missing hit. With this chickpea-specific model, a total of 201 and 256 NBS-candidate proteins were identified in the two varieties. Out of these, 104 and 119 sequences, identified by above two methods, were selected as true NBS-LRR candidate proteins. The gene identification strategy followed in this study is shown in Supplementary Fig. S2.

From the broadly classified eight groups of plant resistance genes based on the motif organization and spanning regions (1) NBS-LRR-TIR, (2) NBS-LRR-CC, (3) LRR-TrDD, (4) LRR-TrDDKINASE, (5) TrDD-CC, (6) LRR-TrDD-PEST-ECs, (7) TIR-NBS-LRR-NLS-WRKY, (8) KINASE-KINASE-KINASE-HM1)) [35], the seventh one with NBS domain (TIR-NBS-LRR-NLS-WRKY) was not found in the chickpea genome.

Varshney et al. [36] reported presence of 187 disease resistance gene homologs (RGHs) in kabuli chickpea variety. Parween et al. [37] identified 133 RGHs in desi chickpea. The genome-wide study of NBS-LRR genes in fabaceae family by Zhang et al. [31], identified 227 R-genes in chickpea. However, we could identify only 104 and 119 NBS-LRR disease resistance genes in the two varieties. There might be a possibility that the extra sequences identified as RGHs in the above three studies does not have NBS domain and belong to other classes of RGs mentioned above with no NBS domain [35]. In order to check this, we analyzed the domain architecture of NBS-LRR protein dataset of Zhang et al. [31]. The additional proteins, identified as NBS-LRR proteins in this study, consist of only “LRR”, “TIR”, “CC”, “No domain”, “RNI-like” domains but no NB-ARC domain. To further validate this hypothesis, we looked for the domains present in the sequences, which are excluded from our analysis. Most of the excluded hits in kabuli chickpea (68 out of 172) belonged to AAA family of ATPase (20/68) and ABC transporter Pfam family (26/68). We could also identify few proteins with kinase domain (PRK) with multiple LRR domains and few are those which only posses TIR/CC domain without NB-ARC domain.

2.2. Orthologs identification

Out of 104 kabuli NBS-LRR genes, we could find orthologs for 100 genes in desi chickpea. Moreover, we observed that > 75% of the total NBS-LRR genes in one variety were similar to the NBS-LRR genes of other variety (> 90% identity) and resides in the syntenic regions (Supplementary Fig. S3). Therefore in the further study, we have analyzed the NBS-LRR genes identified in kabuli variety.

The NBS-LRR genes from other closely related fabaceae species were obtained from previously published studies [19,38,39]. Due to unavailability of published dataset of NBS-LRR genes for C. cajan genome, we identified this gene family in C. cajan using the gene identification approach mentioned in the paper. For the closely related fabaceae species, G. max (36 NBS-LRR orthologs), M. truncatula (72 NBS-LRR orthologs), L. japonicus (26 NBS-LRR orthologs), and C. cajan (50 NBS-LRR genes), orthologs for ≥ 50% chickpea NBS-LRR genes were found. The count of NBS-LRR orthologs genes in respective species was found to be correlated with the total number of NBS genes present in the species and evolutionary distance from kabuli variety (Supplementary Table S2).

2.3. Comparative analysis of kabuli chickpea with desi chickpea and M. truncatula based on homology and synteny of NBS-LRR genes

Syntenic genomic regions for kabuli variety (with desi variety and M. truncatula) were obtained from https://legumeinfo.org. Using above approach, 25,823 and 20,036 pairs of orthologous genes were identified. Out of 104 NBS-LRR genes, we could recover orthologs for 100 and 72 NBS-LRR genes in ICC4958 and M. truncatula respectively. For synteny analysis of NBS-LRR genes, we included ortholog gene pairs anchored on eight major chromosomes in both species. 38 ICC4958 NBS-LRR genes and 43 M. truncatula NBS-LRR genes were found in macro-syntenic blocks (https://legumeinfo.org/data/public/Cicer_arietinum/CDCFrontier.gnm1.snt11/) (Supplementary Fig. S3, Fig. 1). This high fraction of syntenic NBS-LRR genes underlies positional conservation along with sequence conservation in fabaceae species. It also suggests that these genes arose prior to divergence of lineages that led to chickpea and M. truncatula.

2.4. Distribution and clustering of NBS-LRR genes

Some of the NBS-LRR resistance genes are present on chromosomes in isolation whereas others are part of multi-gene clusters. The numbers per chromosome of 87 NBS-LRR resistance genes in chickpea genome distributed in Chromosomes 1 to 8 is 10, 11, 12, 9, 16, 11, 12 and 6 (Supplementary Fig. S1; Supplementary Table S3, S4, S5). Mapping the remaining 17 genes on genome could not be accomplished, as they are located on different scaffolds.

A gene cluster is defined if two neighboring homologous genes are < 200 kb apart and contain < 8 non-NBS resistance genes between two NBS resistance genes [17,25,40]. Moreover, populations from a common ancestor tend to possess the same set of gene clusters that help to trace their recent evolutionary history. There are 21 gene clusters comprising of 49 NBS-LRR resistance genes in chickpea. Among these 21 clusters, four clusters are located on chromosomes 5, three on chromosomes 1, 2, 4, and 7 and two on chromosomes 6 and 8. Only one cluster was observed on chromosome 3 and scaffold 242. Most of the gene clusters have two genes except those on chromosomes 1, 3, 4 and 5.
5, which contained 3 to 4 genes (Supplementary Table S6).

The degree of clustering is comparatively lower in chickpea (47% of the total NBS genes) as compared to other species with a very low number of NBS-LRR genes for instance, cucumber (57 NBS-LRR genes; 58% of the total NBS-LRR genes are clustered). The largest cluster in chickpea comprises of 4 genes. However, largest cluster in cucumber consists of 10 genes. Similar approach resulted in high degree of clustering in other species like \textit{M. domestica} (751 genes in 159 clusters), \textit{Cassava} (143 genes in 39 clusters), and \textit{E. grandis} (1130 genes in 136 clusters).

2.5. Motif identification

For identifying the degree of conservation and presence of signature motifs in protein families of NBS-LRR type, motif analysis was carried out in MEME suite. The NBS-LRR proteins generally have an N-terminal region (CC or TIR), NBS domain and LRR regions. For the sake of motif analysis, protein sequences in the two families are divided into three parts i.e. non-TIR/TIR, NBS, and LRR.

2.5.1. Non-TIR-NBS-LRR family

The types of motifs identified were two CC, ten NBS and seven LRR in the sequences analyzed here. The presence of two CC motifs was common among the members of the non-TNL family. Also, most of the sites in these two motifs were weakly conserved. The NBS domain, on the other hand, showed a higher degree of conservation. The seven conserved motifs (P-loop, RNBS-A-nonTIR, Kinase-2, RNBS-B, RNBS-C, GLPL and MHDL) occurred in most of the non-TNL proteins. The RNBS-D-nonTIR motif was present in only a few sites. Apart from that, two additional motifs CNBS1 and CNBS2 were present in the NBS domain of non-TNL proteins. A similar pattern was seen in cucumber genome as well [22]. Most of the sites in these two motifs were poorly conserved.

Seven LRR motifs were also found in non-TNL proteins. The pattern of occurrence of LRR motifs in these proteins was highly variable. A majority of them showed different LRR motifs. The motifs L1, L2, L3, and L7 were widely present in most of the non-TNL proteins (Supplementary Table S7, Supplementary Table S8, Supplementary Fig. S4). The NBS proteins in cucumber have three CC motifs, however, in chickpea, we could identify only two such motifs.

2.5.2. TIR-NBS-LRR family

A total of four, nine and six motifs were identified in N-terminal (TIR), NBS and LRR domains, respectively, of TNL family. Compared to non-TNL proteins, the conservation level of the NBS domain was high in TNL members (Supplementary Fig. S5, Supplementary Fig. S6). All the nine motifs existed widely in more than half of the TNL proteins. Just like in non-TNL family, one additional motif, namely TNBS-1, is present in the members of this family.

Most of the sites in LRR motifs were weakly conserved. The LRR motifs in TNL family were different from the one observed in non-TNL family (Supplementary Table S9, Supplementary Table S10, Supplementary Fig. S7). Six conserved NBS motifs were present in both NBS resistance gene families (P-loop, Kinase-2, RNBS-B, RNBS-C, GLPL, and MHDL). However, RNBS-A-nonTIR and RNBS-D-nonTIR motifs were observed in only non-TNL members while RNBS-A-TIR and RNBS-D-TIR motifs were present in TNL proteins.

Three additional motifs CNBS1, CNBS2 and TNBS1 were observed in chickpea, which were not found even in its closest relative \textit{M. truncatula}.
[19]. These three NBS motifs were also located in NBS resistance genes of cucumber [22]. These motifs unique to chickpea and cucumber also distinguish the two families of NBS genes.

We also analyzed those motifs where the frequency of occurrence was > 80% so as to show only significant motifs. In non-TNL family, we observed that out of 20 motifs, frequency of occurrence of 11 motifs is more than 80% (P-loop, RNBS-A-nonTIR, Kinase-2, RNBS-B, RNBS-C, GLPL, MHDIL, L1, L2, L3, and L7). Similarly, 7 out of 20 motifs occurred > 80% in TNL family also (P-loop, Kinase-2, RNBS-B, RNBS-C, GLPL, TNBS-1, MHDIL).

2.6. Phylogenetic analysis and gene duplication

Generally, the 5′ region preceding and the 3′ region following the NBS domain have a high degree of sequence variability and therefore not considered for building the phylogenies. Meyers et al. [41] reported that the phylogenetic analysis considering NBS domain classifies the sequences into non-TNL and TNL families. Therefore, NBS domains of these 104 proteins (P-loop to GLPL) were extracted, based on the motifs identified by MEME, for constructing the Neighbor Joining (NJ) phylogenetic tree (Supplementary Fig. S5, Supplementary Fig. S6). We observed that out of 104, 93 NBS-LRR proteins have complete NBS domain. The remaining 11 sequences were not considered for the phylogenetic analysis because either their NBS domains were incomplete or the signature motifs of NBS domain were less conserved.

The two families distinctly formed two separate clades in the dendrogram with high bootstrap values (Fig. 2). The number of non-TNL proteins was more than the number of TNL ones consisting of 81 and 23 members, respectively. This finding is in agreement with the distribution of NBS-LRR resistance genes in cucumber genome [22], however, different from the distribution in A. thaliana and E. grandis where the members of TNL outnumber non-TNL genes [17,42].

In total, 74 recent gene duplication events, comprising of 40 genes, were seen in NBS-LRR gene family in chickpea. Chromosomes 1 and 5 harbors most of the duplicated genes (30 and 14 genes out of 74 genes). Gene duplication contributes to high number and diversity of NBS-LRR genes in other species for instance 174, 519, 416 NBS genes in Arabidopsis, rice, and poplar respectively.

2.7. Domain distribution and arrangement

Domain arrangements in the 104 NBS-LRR proteins of chickpea were analyzed using hidden Markov model (HMM) search against Pfam database. Pfam database does not predict CC motifs in the N-terminal region. Previous studies have suggested that the presence of specific signature motifs in NBS domain can be correlated with the presence or absence of CC motifs [41,43]. We have used these signature sequences to classify the non-TNL family. The presence of CC regions was further validated by HMMER search (phmmer) against UniProKB protein database and MARCOIL server [44] using 9FAM matrix. The Probability of N-terminal region showing Coiled Coil conformation in the MARCOIL plot was in the range of 0.4-1 for most of the non-TNL proteins.

The canonical form of the domain (CNL & TNL) was observed in 19 non-TNL and 6 TNL proteins. In the non-TNL family, the predominant domain arrangement was CN type in 33 of the total 77 with missing LRR domain similar to other plant genomes like M. truncatula [19]. Another class of domains in some non-TNL proteins was the result of RPW8 domain fusing with N or CNL domain seen in Ca_10860, Ca_10861, and Ca_02119. The RPW8 gene in A. thaliana provides example of a broad-spectrum resistance genes against powdery mildew [45]. Compared to non-TNL, a more diverse arrangement of domains was found in TNL family namely TNL, TN, NL, N, L, TNTN, NN, TTNL, and NTN (Supplementary Table S4). Apart from the domain shuffling explained by Meyers et al. 2002, four new domains- CNLNI, CNNL, TNTN, and NTN, were also identified here in these two families (Supplementary Table S11).

Some of the TNL proteins in M. truncatula possess > 1 NBS domain (NTNL, TNTNL, TNTNL). We observed four chickpea NBS proteins depicting the similar domain structure (NTNL, TNTNL, NNLNL, TNTL) with only one complete NBS domain and other one truncated.

2.8. Exon-intron architecture

The detailed illustration of exon-intron arrangement in chickpea NBS-LRR resistance genes is shown in Supplementary Fig. S8. The number of introns varies from 0 to 11 in both the families. Interestingly, no TNL gene was without an intron, whereas 25 non-TNL genes had no introns. More than half of the sequences (52 out of 81) had either no introns or up to 3 introns (Supplementary Table S3, S4). The least number of introns (1–4) was present in all members of TNL family except Ca_10064, which contained 11 introns (Supplementary Table S3, S4). These findings suggest that the intron gain and loss events may be progressively occurring during the structural evolution of the two families of chickpea NBS resistance genes.

2.9. Gene expression analysis through RNA-seq

Out of the 104 genes identified, 79 showed medium to high expression (FPKM ≥ 5) in at least one of the 5 tissues selected (Flower, Flower bud, shoot apical meristem, young leaves and germinating seedling) whereas 10 showed low expression (5 > FPKM > 0) and 15 showed no expression (FPKM = 0) in all the five tissues under study (Supplementary Table S12). Differential expression patterns were observed across the tissues with most of the non-TNL and TNL genes showing highest expression in germinating seedling and shoot apical meristem (Fig. 3).

2.10. In silico promoter analysis of NBS-LRR resistance genes

A 2 kb upstream region of the NBS-LRR resistance genes was searched for regulatory elements. Four cis-regulatory elements related to stress conditions and pathogen attack were found overrepresented in the 2 kb upstream region of NBS resistance genes. The four promoter elements considered were WBOX associated with WRKY transcription factor [46], DRE [47], CBF [48] and GCC box. WBOX elements were
M. Table S3, S4). A similar pattern of occur only once with few exceptions that occurs twice (Supplementary The occurrence of two WBOXs was common in 23 NBS-LRR resistance quite a low number averaging 0.20 (DRE), 0.10 (CBF) and 0.12 (GCC). Ca_13840, which had no WBOX. Rest of the three boxes were present in 4.52 for TNL excluding few like Ca_08949, Ca_13839, Ca_08628, and Palomino et al. [49] and Huettel et al. [50]. Similar identi methods. This includes 25 and 8 NBS-LRRs previously reported by proteome, using the chickpea proteome by iterative computational kabuli variety of chickpea, which constitute about 0.36% of the total present study, we report the presence of 104 NBS resistance genes in various plant genomes as mentioned in the introduction section. In the distribution of non-TNL and TNL in chickpea genome may suggest a ratio of 1:2 (CNL: TNL) has been reported in the of 81 non-TNL and 23 TNL, making a ratio of approximately 4:1. While, The NBS encoding genes were classi into two broad groups i.e. TNL and non-TNL based on the amino-terminal region and the motifs. Our analysis supported the existence of distinction among the TNL/non-TNL groups. NBS encoding genes in chickpea constitute a total number of 81 non-TNL and 23 TNL, making a ratio of approximately 4:1. While, A. thaliana, A. lyrata and Brassica rapa [17,23,51] and a ratio of 4:1 (CNL: TNL) observed in potato [40] and grapevine genome [25]. The distribution of non-TNL and TNL in chickpea genome may suggest higher contribution of non-TNL in response to the pathogen attack. Although the number of NBS resistance genes in chickpea is quite less as compared to some other sequenced genomes, it has both the families TNL and non-TNL, which suggests that chickpea has few but diverse set of resistance genes. The NBS motifs also showed diversity in the two families. Six motifs (P-loop, Kinase-2, RNBS-B, RNBS-C, GLPL, and MHDL) are common to both families, whereas RNBS-A-TIR and RNBS-D-TIR are exclusive to TIR-NBS family. Similarly, RNBS-A-
non-TIR and RNBS-D-nonTIR are specific to non-TIR-NBS family. The position of tryptophan or aspartic acid residues of the kinase-2 motif (DDWV/DDVD) distinguishes TNL from non-TNL family with 95% accuracy. In chickpea, the last amino acid residue is tryptophan in all non-TNL proteins except Ca.23573, Ca.21455, Ca.08360 and Ca.08364 in which it is replaced by arginine, cysteine, and glutamic acid respectively. Although 11 TNL proteins have aspartic acid as the last residue of kinase-2 motif, it is replaced by glutamic acid, alanine, serine, glycine or asparagine in remaining TNL sequences.

The expression pattern of NBS-LRR genes in chickpea was analyzed using publicly available RNA-seq data. The expression data showed that 10 NBS resistance genes expressed at low level and fifteen genes remained unexpressed in all the five tissues taken for the study. This observation is supported by expression pattern of NBS resistance genes in Arabidopsis in which the expression has been at low levels and with a variety of tissue specificities [18]. In addition to that, the tissue specific expression profiling will aid in identifying the decisive role of these genes in different plant tissues in conferring the resistance against pathogens and pests. Importantly, pathogen-responsive NBS-LRR genes identified in the current study may be used as candidate genes for enhancing pathogen resistance in chickpea and in other related species too.

4. Methods

4.1. Identification of genes for NBS-LRR proteins in chickpea

The draft genome of chickpea was downloaded from Legume information system (LIS, https://legumeinfo.org/). Genome sequences of two different varieties of chickpea, namely CDC frontier (kabuli, estimated genome size 740 Mb) and ICC4958 (desi, estimated genome size 1.01 Gb), are available in this database. These genomes are annotated with 28,269 protein-coding gene models in kabuli and 30,868 gene models in desi varieties. Pertaining to evolutionary closeness, our study is mainly focused on kabuli variety with comparative analysis with NBS-LRR genes of desi variety. The draft assembly of kabuli variety is distributed over 7163 scaffolds covering 544.73 Mb genome (over 70% of the estimated genome size).

The NBS-LRR proteins were identified in the predicted chickpea proteome using Basic Local Alignment Search Tool (BLAST) (stand-alone blastp 2.2.22) [52] with an E-value cut-off of 10−4. Consensus TNL and non-TNL sequences from plant extended NBS domain defined by Cannon et al. were used in blast query [53]. The resultant hits were further verified by searching NB-ARC domain (Pfam Pf00931) in the proteome of chickpea using the hidden Markov model (HMM) profile of Pfam 27.0 [54] in HMMER 3.0 [55]. The E-value cut-off used was 10−4. From the resultant hits, the high quality ones were used to construct a chickpea-specific NBS hmm-profile using the module “hmmbuild” to check for any missing hit.

4.2. Phylogenetic analysis and gene duplication

The dendrogram was constructed from the conserved NBS motifs (P-loop to GLPL) of NBS proteins (sequence length ≥ 150 amino acid). The sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE) by selecting Unweighted Pair Group Method with Arithmetic Mean (UPGMA) based clustering [56]. Gap open and gap extension penalties were set to −2.9 and 0. This sequence alignment was used to generate the dendrogram using the Neighbor Joining (NJ) algorithm and Dayhoff substitution matrix (PAM250) implemented in Molecular Evolutionary Genetics Analysis 5.1 (MEGA) [57]. A bootstrap value of 1000 replicates was selected to consolidate the grouping pattern in the dendrogram. Gene duplication events in NBS-encoding gene family were defined based on the criteria considered by the previous study by Cheng et al. [20]. This study defines a duplication event when (a) the alignment covered > 70% of the longer gene; (b) the aligned region had an identity > 70%.

4.3. Search for orthologs in other leguminous plants

Orthologs for chickpea NBS-LRR proteins were searched in four fabaceae plant genomes namely, M. truncatula, G. max, C. cajan and L. japonicus using reciprocal best hit approach keeping the sequence identity cut-off ≥ 80% (over at least 50% of query sequence) and E-value cut-off 10−3. These plants were selected based on the evolutionary closeness to C. arietinum reported by Varshney et al. [36].

4.4. Domain classification and motif identification

Domain architecture of the NBS-LRR proteins was analyzed using hmmscan search in HMMER against the Pfam database using gathering threshold. Sequence motifs were predicted using MEME suite with minimum and maximum width of the motif set to 6 and 20 in order to search for a maximum of 20 motifs with zoops model [58,59].

4.5. Exon-intron architecture

Information on gene architecture and exon-intron position of NBS-LRR resistance genes in chickpea was gathered and analyzed using Gene Structure Display Server [60] by comparing gene sequences and coding sequences.

4.6. Gene expression studies using RNA-seq data

RNA-seq raw read data for 5 different tissues of ICC4958 namely germinating seedling (GSM1047862), young leaves (GSM1047863), shoot apical meristem (GSM1047864), flower bud (GSM1047865, GSM1047866, GSM1047867, GSM1047868) and flower (GSM1047869, GSM1047870, GSM1047871, GSM1047872) from Sequence Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) in NCBI database [61]. Reads were quality trimmed using Trim Galore [parameter: -q 26 –fastqc –a], a wrapper tool around Cutadapt [62] and FastQC [63]. Reads were mapped to the genomic sequence of C. arietinum with TopHat, a spliced read mapper [64]. Cufflinks tool was utilized to estimate abundance of reads mapped to genes body [65]. It calculates Fragments Per Kilobase of transcript per Million (FPKM) values as a proxy for gene expression in different tissues.

4.7. In silico promoter analysis

The 2 kb upstream regions of the NBS-LRR resistance genes were selected and then screened against the PLACE (Plant cis-acting regulatory DNA elements) database to identify motifs found in cis-acting regulatory DNA elements [66]. The overrepresented cis-regulatory elements that are also known to be involved in regulation during stress conditions and disease resistance response were selected for analysis. The four important regulatory elements, WBOX [TGAC(C/T)], CBF [GTCGAC], DRE [(G/A) CGGAC] and GCC boxes, which are enriched in the close relative of chickpea i.e., M. truncatula were probed.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2017.08.004.

Conflict of interest

The authors declare that they have no competing interests.

Acknowledgements

RS thanks Council of Scientific and Industrial Research (CSIR), India for Senior Research Fellowship. This work is carried out under CSIR-National Chemical Laboratory Centre of Excellence in Scientific Computing (CoESC) project.


