Genomics Data 14 (2017) 24–31

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

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

Ranu Sharma[a,*,1], Vimal Rawat[b,1], C.G. Suresh[a]

[a] Division of Biochemical Sciences, CSIR-National Chemical Laboratory, Pune, Maharashtra, India
[b] Institute of Plant and Microbial Biology and Zurich-Basel Plant Science Center, University of Zurich, Zollikerstrasse, 107, CH 8008, Zurich, Switzerland

ARTICLE INFO

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

ABSTRACT

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 specificity [3,4]. Based on the structure of the N-terminal domain, the NBS-LRR genes are divided into two families. The N-terminal domain of one of the families possesses homology with drosophila Toll and human interleukin-1 receptors (TIR) therefore known as TIR-NBS-LRR (TNL), which is known to be involved in resistance specificity and signaling [5,6]. The other family, where the TIR is absent or in its place a coiled-coil (CC) N-terminal domain involved in protein-protein interactions 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).
[1] 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

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 (*Aschochyta 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 the 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 (hmmsearch) 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 membrane spanning regions [(1) NBS-LRR-TIR, (2) NBS-LRR-CC, (3) LRR-TrD, (4) LRR-TrD-KINASE, (5) TrD-CC, (6) LRR-TrD-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 RGSs 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.synt1/) (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

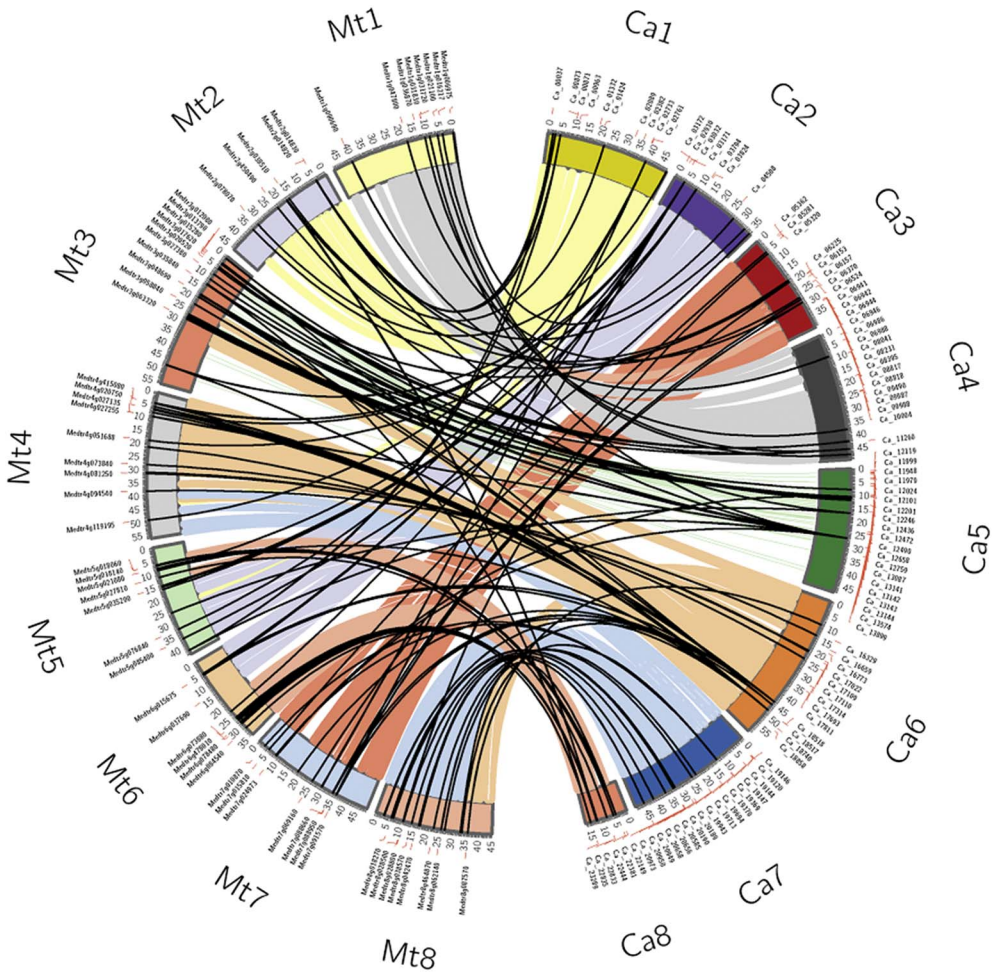

Fig. 1. Distribution and synteny analysis of NBS-LRR genes on kabuli (Cal–Ca8) and *M. truncatula* (Mt1–Mt8) chromosomes. NBS-LRR genes are indicated by vertical black lines. Colored bands denote syntenic regions of the two genomes.

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 *M. domesticus* (751 genes in 159 clusters), Cassava (143 genes in 39 clusters), and *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 *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 that 80% (P-loop, RNBS-A-nonTIR, Kinase-2, RNBS-B, RNBS-C, GLPL, MHDL, 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, MHDL).

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.

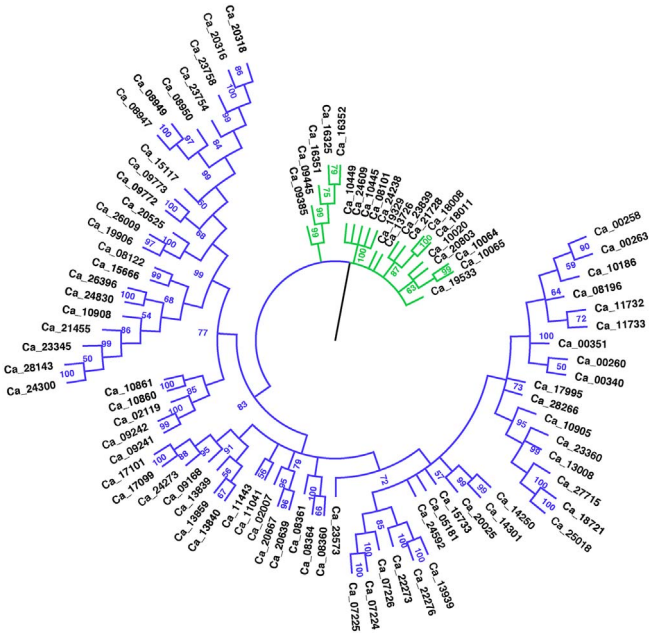

Fig. 2. Phylogenetic analysis of NBS-LRR proteins. Circular representation of dendrogram reveals distinct clusters of non-TNL (blue clad) and TNL (green clad) chickpea proteins.

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 UniProtKB 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- CNNL, CNN, 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, TNLTNL). We observed four chickpea NBS proteins depicting the similar domain structure (NTNL, TNTNL, NLNL, TTNL) 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

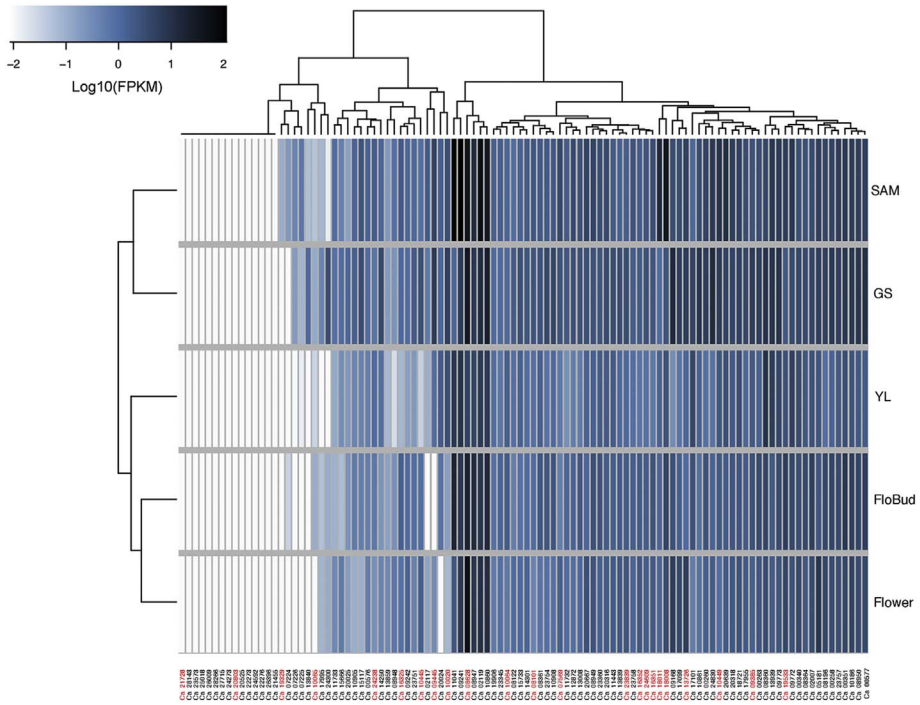

Fig. 3. Expression level of chickpea NBS-LRR genes. The heatmap depicts relative gene expression of non-TNL (black) and TNL (red) genes in various tissue samples (Shoot apical meristem (SAM), Germinating seedling (GS), Young leaves (YL), Flower bud (FloBud), Flower). The color scale (− 2 to 2) represents $\log_{10}$(FPKM), calculated by comparing Fragments Per Kilobase of transcript per Million (FPKM) value for NBS-LRR genes in different tissues. Dendrogram on the top and side of the heatmap shows hierarchical clustering of tissues and genes using complete linkage approach.

widely present in both the families averaging at 3.6 for non-TNL and 4.52 for TNL excluding few like Ca_08949, Ca_13839, Ca_08628, and Ca_13840, which had no WBOX. Rest of the three boxes were present in quite a low number averaging 0.20 (DRE), 0.10 (CBF) and 0.12 (GCC). The occurrence of two WBOXs was common in 23 NBS-LRR resistance genes, nine WBOXs being the maximum. The other regulatory elements occur only once with few exceptions that occurs twice (Supplementary Table S3, S4). A similar pattern of *cis*-regulatory elements occurring in *M. truncatula* was observed [19].

3. Discussion

Chickpea is an economically important legume crop widely consumed all over the world owing to its rich protein content and high nutrition value. Diseases such as Ascochyta blight and Fusarium wilt are some of the root causes of its reduced productivity. In order to reduce crop loss due to these diseases, understanding and enhancing disease resistance is very crucial. The NBS encoding disease resistance genes play an important role in protecting plants from diverse range of pathogens and insect pests. Availability of sequenced chickpea genome makes it possible to carry out genomic studies on NBS genes that confer resistance to rapidly evolving pathogens. A deeper understanding of location, structure and gene expression of disease resistance genes can assist the plant breeders to improve disease resistance capacity of the chickpea crop.

The NBS disease resistance genes have been studied extensively in various plant genomes as mentioned in the introduction section. In the present study, we report the presence of 104 NBS resistance genes in kabuli variety of chickpea, which constitute about 0.36% of the total proteome, using the chickpea proteome by iterative computational methods. This includes 25 and 8 NBS-LRRs previously reported by Palomino et al. [49] and Huettel et al. [50]. Similar identification studies on *M. truncatula*, *P. trichocarpa*, and *A. thaliana* genomes revealed presence of 333, 402, and 207 NBS resistance genes. Thus, chickpea seems to encode comparatively lesser number of NBS resistance genes. However, a similar pattern of lesser number of NBS resistance genes was also observed in plants such as *B. rapa* (92), *Z. mays* (109), cucumber (57), and *Carica papaya* (54). Wan et al. [22] explained the presence of low number of NBS resistance genes in cucumber to be due to absence of whole-genome duplication, small number of tandem gene duplication and few segmental duplication. Fewer duplication events might be the reason for the reduced number of NBS resistance genes in chickpea too.

In rice, 70% of the total NBS resistance genes occurred in 104 gene clusters [25] whereas in chickpea only 48% of the total NBS genes present in only 21 gene clusters. Another reason behind low NBS gene count in chickpea could be reduced clustered arrangement of this gene family. The complex clustering of NBS resistance genes greatly contributes to genetic variations [22]. This would have probably influenced the total number NBS resistance genes present in chickpea.

Previous gene identification studies by Huettel et al. [50], Palomino et al. [49], and Zhang et al. [31] were based on NB-ARC domain identification and blast search for identifying candidate genes by sharing significant protein similarity with known plant resistance gene analogs (RGAs) in protein database. However, we used three different methods, based on blastp search, NB-ARC domain search and search using chickpea-specific NBS hidden Markov model; all the three methods confirmed the presence of same number of NBS proteins (104) in chickpea. In addition, we observed the presence of other RGHs in chickpea that also contribute in providing resistance to pest and pathogen attack.

The NBS encoding genes were classified 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 ratio of 1:2 (CNL: TNL) has been reported in the *brassicaeae* family including *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-

nonTIR and RNBS-D-nonTIR are specific to non-TIR-NBS family. The position of tryptophan or aspartic acid residues of the kinase-2 motif (DDVW/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 30,686 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) (standalone 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) CCGAC] 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.

References

[1] X. Dong, Sa, ja, ethylene, and disease resistance in plants, Curr. Opin. Plant Biol. 1 (1998) 316–323.

[2] W.I.L. Tameling, et al., The tomato r gene products i-2 and mi-1 are functional atp binding proteins with atpase activity, Plant Cell 14 (2002) 2929–2939.

[3] B. Kobe, J. Deisenhofer, A structural basis of the interactions between leucine-rich repeats and protein ligands, Nature 374 (1995) 183–186, http://dx.doi.org/10.1038/374183a0.

[4] R.T. Leister, F. Katagiri, A resistance gene product of the nucleotide binding site – leucine rich repeats class can form a complex with bacterial avirulence proteins in vivo, Plant J. 22 (2000) 345–354.

[5] B.J. DeYoung, R.W. Innes, Plant nbs-lrr proteins in pathogen sensing and host defense, Nat. Immunol. 7 (2006) 1243–1249, http://dx.doi.org/10.1038/ni1410.

[6] J.E. Luck, G.J. Lawrence, P.N. Dodds, K.W. Shepherd, J.G. Ellis, Regions outside of the leucine-rich repeats of flax rust resistance proteins play a role in specificity determination, Plant Cell 12 (2000) 1367–1377.

[7] G.B. Martin, A.J. Bogdanove, G. Sessa, Understanding the functions of plant disease resistance proteins, Annu. Rev. Plant Biol. 54 (2003) 23–61, http://dx.doi.org/10.1146/annurev.arplant.54.031902.135035.

[8] G. van Ooijen, H.A. van den Burg, B.J.C. Cornelissen, F.L.W. Takken, Structure and function of resistance proteins in solanaceous plants, Annu. Rev. Phytopathol. 45 (2007) 43–72, http://dx.doi.org/10.1146/annurev.phyto.45.062806.094430.

[9] N. Aarts, et al., Different requirements for eds1 and ndr1 by disease resistance genes define at least two r gene-mediated signaling pathways in arabidopsis, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 10306–10311.

[10] H. Zhu, S.B. Cannon, N.D. Young, D.R. Cook, Phylogeny and genomic organization of the tir and non-tir nbs-lrr resistance gene family in medicago truncatula, Mol. Plant-Microbe Interact. 15 (2002) 529–539, http://dx.doi.org/10.1094/MPMI.2002.15.6.529.

[11] M.W.F. Yaish, L.E. Sáenz de Miera, M. Pérez de la Vega, Isolation of a family of resistance gene analogue sequences of the nucleotide binding site (nbs) type from lens species, Genome 47 (2004) 650–659, http://dx.doi.org/10.1139/g04-027.

[12] J.-X. Yue, B.C. Meyers, J.-Q. Chen, D. Tian, S. Yang, Tracing the origin and evolutionary history of plant nucleotide- binding site-leucine-rich repeat (nbs-lrr) genes, New Phytol. 193 (2012) 1049–1063, http://dx.doi.org/10.1111/j.1469-8137.2011.04006.x.

[13] B. Monosi, R.J. Wisser, L. Pennill, S.H. Hulbert, Full-genome analysis of resistance gene homologues in rice, TAG Theor. Appl. Genet. 109 (2004) 1434–1447, http://dx.doi.org/10.1007/s00122-004-1758-x.

[14] T. Zhou, et al., Genome-wide identification of nbs genes in japonica rice reveals significant expansion of divergent non-tir nbs-lrr genes, Mol. Gen. Genomics. 271 (2004) 402–415, http://dx.doi.org/10.1007/s00438-004-0990-z.

[15] P. Arya, G. Kumar, V. Acharya, A.K. Singh, Genome-wide identification and expression analysis of nbs-encoding genes in malus x domestica and expansion of nbs genes family in rosaceae, PLoS One 9 (2014) e107987, http://dx.doi.org/10.1371/jour- nal.pone.0107987.

[16] M. Perazzolli, et al., Characterization of resistance gene analogues (rgas) in apple (malus ∗ domestica borkh.) and their evolutionary history of the rosaceae family, PLoS One 9 (e83844) (2014), http://dx.doi.org/10.1371/journal.pone.0083844.

[17] B.C. Meyers, A. Kozik, A. Griego, H. Kuang, R.W. Michelmore, Genome-wide analysis of nbs-lrr-encoding genes in arabidopsis, Plant Cell 15 (2003) 809–834.

[18] X. Tan, et al., Global expression analysis of nucleotide binding site-leucine rich repeat-encoding and related genes in arabidopsis, BMC Plant Biol. 7 (2007) 56, http://dx.doi.org/10.1186/1471-2229-7-56.

[19] C. Ameline-Torregrosa, et al., Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant medicago truncatula, Plant Physiol. 146 (2008) 5–21, http://dx.doi.org/10.1104/pp.107.104588.

[20] Y. Cheng, et al., Systematic analysis and comparison of nucleotide-binding site disease resistance genes in maize, FEBS J. 279 (2012) 2431–2443, http://dx.doi.org/10.1111/j.1742-4658.2012.08621.x.

[21] B.W. Porter, et al., Genome-wide analysis of carica papaya reveals a small nbs resistance gene family, Mol. Gen. Genomics. 281 (2009) 609–626, http://dx.doi.org/10.1007/s00438-009-0434-x.

[22] H. Wan, et al., Genome-wide analysis of nbs-encoding disease resistance genes in cucumis sativus and phylogenetic study of nbs-encoding genes in cucurbitaceae crops, BMC Genomics 14 (2013) 109, http://dx.doi.org/10.1186/1471-2164-14-109.

[23] J.-H. Mun, H.-J. Yu, S. Park, B.-S. Park, Genome-wide identification of nbs-encoding resistance genes in brassica rapa, Mol. Gen. Genomics. 282 (2009) 617–631, http://dx.doi.org/10.1007/s00438-009-0492-0.

[24] A. Kohler, et al., Genome-wide identification of nbs resistance genes in populus trichocarpa, Plant Mol. Biol. 66 (2008) 619–636, http://dx.doi.org/10.1007/s11103-008-9293-9.

[25] S. Yang, X. Zhang, J.-X. Yue, D. Tian, J.-Q. Chen, Recent duplications dominate nbs-encoding gene expansion in two woody species, Mol. Gen. Genomics. 280 (2008) 187–198, http://dx.doi.org/10.1007/s00438-008-0355-0.

[26] R. Lozano, O. Ponce, M. Ramirez, N. Mostajo, G. Orjeda, Genome-wide identification and mapping of nbs-encoding resistance genes in solanum tuberosum group phureja, PLoS One 7 (2012) e34775, http://dx.doi.org/10.1371/journal.pone.0034775.

[27] S.M. Kale, V.C. Pardeshi, V.T. Barvkar, V.S. Gupta, N.Y. Kadoo, Genome-wide identification and characterization of nucleotide binding site leucine-rich repeat genes in linseed reveal distinct patterns of gene structure, Genome 56 (2013) 91–99, http://dx.doi.org/10.1139/gen-2012-0135.

[28] H. Wei, W. Li, X. Sun, S. Zhu, J. Zhu, Systematic analysis and comparison of nucleotide-binding site disease resistance genes in a diploid cotton gossypium raimondii, PLoS One 8 (2013) e68435, http://dx.doi.org/10.1371/journal.pone.0068435.

[29] H. Song, et al., Comparative analysis of nbs-lrr genes and their response to aspergillus flavus in arachis, PLoS One 12 (2017) e0171181, http://dx.doi.org/10.1371/journal.pone.0171181.

[30] Y. Li, Y. Zhong, K. Huang, Z.-M. Cheng, Genomewide analysis of nbs-encoding genes in kiwi fruit (actinidia chinensis), J. Genet. 95 (2016) 997–1001.

[31] Y. Zhang, R. Xia, H. Kuang, B.C. Meyers, The diversification of plant nbs-lrr defense genes directs the evolution of micrornas that target them, Mol. Biol. Evol. 33 (2016) 2692–2705, http://dx.doi.org/10.1093/molbev/msw154.

[32] R. Xia, J. Xu, S. Arikit, B.C. Meyers, Extensive families of mirnas and phas loci in norway spruce demonstrate the origins of complex phasirna networks in seed plants, Mol. Biol. Evol. 32 (2015) 2905–2918, http://dx.doi.org/10.1093/mol- bev/msv164.

[33] S. Sharma, N. Yadav, A. Singh, R. Kumar, Nutrition and antinutrition profile of newly developed chickpea (cicer arietinum l) varieties, Int. Food Res. J. 20 (2013) 805–810.

[34] M. Chérif, A. Arfaoui, A. Rhaiem, Phenolic compounds and their role in bio-control and resistance of chickpea to fungal pathogenic attacks, Tunis. J. Plant Prot. 2 (2007) 7–21.

[35] M. Gururani, et al., Plant disease resistance genes: Current status and future directions, Physiol. Mol. Plant Pathol. 78 (2012) 51–65.

[36] R.K. Varshney, et al., Draft genome sequence of chickpea (cicer arietinum) provides a resource for trait improvement, Nat. Biotechnol. 31 (2013) 240–246, http://dx.doi.org/10.1038/nbt.2491.

[37] S. Parween, et al., An advanced draft genome assembly of a desi type chickpea (*cicer arietinum* l.), Sci Rep 5 (12806) (2015), http://dx.doi.org/10.1038/srep12806.

[38] Y.J. Kang, et al., Genome-wide mapping of nbs-lrr genes and their association with disease resistance in soybean, BMC Plant Biol. 12 (2012) 139, http://dx.doi.org/10.1186/1471-2229-12-139.

[39] X. Li, et al., Identification and characterization of nbs-encoding disease resistance genes in lotus japonicus, Plant Syst. Evol. 289 (2010) 101–110.

[40] F. Jupe, et al., Identification and localisation of the nb-lrr gene family within the potato genome, BMC Genomics 13 (2012) 75, http://dx.doi.org/10.1186/1471-2164-13-75.

[41] B.C. Meyers, et al., Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily, Plant J. 20 (1999) 317–332.

[42] N. Christie, P.A. Tobias, S. Naidoo, C. Külheim, The eucalyptus grandis nbs-lrr gene family: Physical clustering and expression hotspots, Front. Plant Sci. 6 (1238) (2015), http://dx.doi.org/10.3389/fpls.2015.01238.

[43] B.C. Meyers, M. Morgante, R.W. Michelmore, Tir-x and tir-nbs proteins: Two new families related to disease resistance tir-nbs-lrr proteins encoded in arabidopsis and other plant genomes, Plant J. 32 (2002) 77–92.

[44] M. Delorenzi, T. Speed, An hmm model for coiled-coil domains and a comparison with pssm-based predictions, Biomaterials 18 (2002) 617–625 (Oxford, England).

[45] S. Xiao, et al., Broad-spectrum mildew resistance in *arabidopsis thaliana* mediated by rpw8, Science 291 (2001) 118–120 (New York, N.Y.), http://dx.doi.org/10.1126/science.291.5501.118.

[46] J. Dong, C. Chen, Z. Chen, Expression profiles of the arabidopsis wrky gene superfamily during plant defense response, Plant Mol. Biol. 51 (2003) 21–37.

[47] Y. Sakuma, et al., Dual function of an arabidopsis transcription factor dreb2a in water-stress-responsive and heat- stress-responsive gene expression, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 18822–18827, http://dx.doi.org/10.1073/pnas.0605639103.

[48] M. Ohme-Takagi, K. Suzuki, H. Shinshi, Regulation of ethylene-induced transcription of defense genes, Plant Cell Physiol. 41 (2000) 1187–1192.

[49] C. Palomino, Z. Satovic, Cubero, J. I. & Torres, A. M. Identification and characterization of nbs-lrr class resistance gene analogs in faba bean (vicia faba l.) and chickpea (cicer arietinum l.), Genome 49 (2006) 1227–1237, http://dx.doi.org/10.1139/g06-071.

[50] B. Huettel, D. Santra, J. Muehlbauer, G. Kahl, Resistance gene analogues of chickpea (*cicer arietinum* l.): isolation, genetic mapping and association with a fusarium resistance gene cluster, TAG Theor. Appl. Genet. 105 (2002) 479–490, http://dx.doi.org/10.1007/s00122-002-0933-1.

[51] Y.-L. Guo, et al., Genome-wide comparison of nucleotide-binding site-leucine-rich repeat-encoding genes in arabidopsis, Plant Physiol. 157 (2011) 757–769, http://dx.doi.org/10.1104/pp.111.181990.

[52] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

[53] S.B. Cannon, et al., Diversity, distribution, and ancient taxonomic relationships within the tir and non-tir nbs-lrr resistance gene subfamilies, J. Mol. Evol. 54 (2002) 548–562, http://dx.doi.org/10.1007/s00239-001-0057-2.

[54] M. Punta, et al., The pfam protein families database, Nucleic Acids Res. 40 (2012) D290–D301, http://dx.doi.org/10.1093/nar/gkr1065.

[55] S.R. Eddy, Profile hidden markov models, Bioinformatics 14 (1998) 755–763 (Oxford, England).

[56] R.C. Edgar, Muscle: Multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (2004) 1792–1797, http://dx.doi.org/10.1093/nar/gkh340.

[57] K. Tamura, et al., Mega5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods, Mol. Biol. Evol. 28 (2011) 2731–2739, http://dx.doi.org/10.1093/molbev/msr121.

[58] T.L. Bailey, et al., Meme suite: Tools for motif discovery and searching, Nucleic Acids Res. 37 (2009) W202–W208, http://dx.doi.org/10.1093/nar/gkp335.
[59] T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, Proc. Int. Conf. on Intell. Syst. Mol. Biol. 2 (1994) 28–36.
[60] A.-Y. Guo, Q.-H. Zhu, X. Chen, J.-C. Luo, [GSDS: a Gene Structure Display Server]. *Yi chuan = Hered*, 29 (2007), pp. 1023–1026.
[61] V.K. Singh, R. Garg, M. Jain, A global view of transcriptome dynamics during flower development in chickpea by deep sequencing, Plant Biotechnol. J. 11 (2013) 691–701, http://dx.doi.org/10.1111/pbi.12059.
[62] M. Martin, Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads, 17 EMBnet.journal, 2010.
[63] S. Andrews, Fastqc: a Quality Control Tool for High Throughput Sequence Data, available online at http://www.bioinformatics.babraham.ac.uk/projects/fastqc, (2010).
[64] C. Trapnell, L. Pachter, S.L. Salzberg, Tophat: discovering splice junctions with rna-seq, Bioinformatics 25 (2009) 1105–1111 (Oxford, England), http://dx.doi.org/10.1093/bioinformatics/btp120.
[65] C. Trapnell, et al., Transcript assembly and quantification by rna-seq reveals unannotated transcripts and isoform switching during cell differentiation, Nat. Biotechnol. 28 (2010) 511–515, http://dx.doi.org/10.1038/nbt.1621.
[66] K. Higo, Y. Ugawa, M. Iwamoto, T. Korenaga, Plant cis-acting regulatory dna elements (place) database: 1999, Nucleic Acids Res. 27 (1999) 297–300.