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