RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2021.054

ScienceAsia 47 (2021): **425–433**

Genome-wide probing of NBS-LRR encoding genes in red clover (*Trifolium pratense* L) for the identification of resistance gene analogs in *Trifolium alexandrinum* L

Bukhtawer Nasir[a], **Siddra Ijaz**[a,*], **Faisal Saeed Awan**[a], **Imran Ul Haq**[b]

[a] Centre of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, University Road, Faisalabad, Pakistan
[b] Department of Plant Pathology, University of Agriculture, University Road, Faisalabad, Pakistan

*Corresponding author, e-mail: siddraijazkhan@yahoo.com

Received 5 Sep 2020
Accepted 16 Apr 2021

ABSTRACT: The nucleotide-binding sites (NBS) domain is the highly conserved domain of nucleotide-binding sites-leucine rich repeats (NBS-LRR) class encoded by resistance genes (*R* genes), and it is used to identify disease resistance-related genes. This study was based on the genome-wide identification of NBS-LRR encoding genes in *Trifolium pratense* to recover resistance gene analogs (RGAs) in *Trifolium alexandrinum*. In *T. pratense*, 251 protein sequences were identified to have the NBS domain. Of these, 16 NBS proteins were predicted to be localized to the mitochondria, 9 to the chloroplasts, 44 to the secretory pathways, and 182 to subcellular locations other than chloroplasts or mitochondria. The structure pattern of predicted NBS genes of the NBS-LRR group was displayed as single gene architecture having an untranslated region (UTRs) with four CD regions. The conserved sequences of the NBS-LRR group were used for primer designing to recover RGAs from *T. alexandrinum*. A maximum likelihood method based phylogram was constructed for exploring the phylogenetic relationship amongst NBS genes encoding protein sequences of *T. pratense* and *R* genes of *Medicago truncatula* and *Medicago sativa* encoding protein sequences having NBS-LRR domain that showed the highly close phylogenetic relationship among them, which supported the highly conserved nature of NBS-LRR class. Hence, we used the conserved NBS-LRR genetic regions to recover the RGAs from the *T. alexandrinum* genome. Therefore, the DNA of a cultivar of *T. alexandrinum* was subjected to polymerase chain reaction (PCR) analysis using the primer designed from these regions; and all primers, except RNL3-F/R, RNL5-F/R, and RNL7-F/R, gave the amplification.

KEYWORDS: nucleotide-binding sites, leucine-rich repeats, *R* genes, resistance gene analogs, *Trifolium*

INTRODUCTION

Plant diseases are a major threat to crop plant health and productivity. Among plant diseases, fungal diseases are the core danger to plants. Upon interaction of fungal pathogens to plants, the physiological and biochemical dynamics of plants are changed adversely. No doubt, plants show their response to those adversities and try to overcome the situation and establish momentum. Even then, the results of that momentum are either positive in terms of resistance, or negative as susceptible [1]. The key players behind the momentum of disease resistance's stability are the genetic elements, which contribute putatively as strong warriors of this battle. Disease resistance genes (*R* genes) are the significant components of resistance mechanisms in plants against pathogens. The exploration of disease resistance mechanisms through traditional and modern approaches is the fundamental task for plant scientists.

Pathogens may also systemically infect the plants by breaking the first line of defense consisting of structural barriers and then enter into plant cells, where they replicate and their infectious proteins start to travel towards other cells [2] invading the second line of defense. As a response to pathogen attack, plants activate defense mechanisms mediated by *R* genes. Gene for gene interaction, which is the interaction between the *avr* gene (avirulence gene) from the plant's pathogen and *R* genes, is activated on their interface [3]. Various resistance genes have been characterized for their resistance to microbial pathogens. Thereby, several resistance gene analogs (RGAs) as candidate *R* genes involving in plant defense against pathogens have been documented.

Sequence analysis of RGAs from different plants characterized to contribute to plant defense mech-

anisms showed significant homology in specific regions, which latterly are known as conserved domains [4–6]. Resistance genes encoding proteins contain conserved domains, i.e., leucine-rich repeats (LRR) and nucleotide-binding site (NBS) domain, and their putative role had been demonstrated in plant disease resistance [7, 8]. The nucleotide-binding site-leucine-rich repeats (NBS-LRR) containing proteins possess a variable N-terminal domain (∼200 amino acids), NBS domain (∼300 amino acids), and LRR domain (∼10–40 short leucine-rich repeats) [9, 10]. Based on the N-terminal region's presence and absence (homology to Toll/ interleukin-1 receptor), plant resistacet genes are subdivided into two families as TIR or non-TIR genes. The non-TIR family has a coiled-coil domain (CC) and leucine zipper (LZ) domain at the N- terminal. The Non-TIR genes are present in both dicot and monocot plant species. However, TIR genes present in only dicotyledonous plants [11, 12]. The NBS domain has many conserved motifs, including the P-loop, kinase-2 domains, kinase-3a domain, and GLPL domain, and a putative membrane-spanning hydrophobic domain [13]. The accessibility to plant genome sequences has impelled scientists to identify and characterize the complete set of NBS-LRR-encoding *R* genes and RGAs in plant species. The molecular biology approaches associated with computational biology have explored several hundred NBS-LRR-encoding genes in *Medicago* species, soybean, alfalfa, chickpea, grapevine [14–16] papaya [17], *Arabidopsis* [18], and rice [19].

By considering the perspectives of *R* genes or RGAs identification and recovery for their possible use as molecular markers for DNA marker-assisted breeding in plants and as transgenes for transgenic plant development as disease management strategies, this study was attempted to identify RGAs in *Trifolium alexandrinum*. Due to the unavailability of the genome sequences of *T. alexandrinum*, genome-wide identification of NBS-LRR encoding RGAs or *R* genes from its closely related species *Trifolium pratense*, whose genome sequence is available, was performed. As discussed, NBS-LRR is a conserved domain, and the conserved motifs of *R* genes offer the opportunity to recover RGAs from other plant species. Hence, using computational biology, NBS-LRR encoding RGAs or *R* genes were identified from *T. pratense*, and *R* genes of cousin species of *T. alexandrinum*, which are *Medicago sativa* and *Medicago truncatula*, were also retrieved from the database. The primers were designed from identified predicted conserved regions to recover RGAs from *T. alexandrinum*.

MATERIALS AND METHODS

In silico mining of the genes encompassing NBS-LRR domain

The genome of *T. alexandrinum* (Berseem/Egyptian clover) is not available; hence, the *T. pratense* (red clover) genome was subjected to genome-wide identification of NBS-LRR encoding *R* genes. Additionally, the *R* genes of *M. sativa*, and *M. truncatula* were retrieved from the plant resistance gene database (http://www.prgdb.crg.eu), as the *R* genes' sequences of Egyptian clover and other *Trifolium* species are not available in the PRG (Plant Resistance Gene) database. *Medicago* species are the cousins of *T. alexandrinum*. They all are members of the family, *Fabaceae*, and the tribe *Trifolieae*. The genome of *T. pratense* (red clover) was accessed from the NCBI database (https://www.ncbi.nlm.nih.gov/nuccore) as well as *Phytozome* (https://phytozome.jgi.doe.gov). The genome of red clover is 309 MB, 40 868 genes, and 42 223 transcripts. Nevertheless, its 164 MB genome is distributed over seven chromosomes having 22 042 genes (https://phytozome.jgi.doe.gov) encoding 64 343 proteins (https://www.ncbi.nlm.nih.gov). InterProScan V. 5.27 and PROSITE ID (https://prosite.expasy.org) were used to find conserved domains such as TIR, CC, RPW8, NBS, and LRR as well as to search the presence of other motifs in these domains. Further validation for the CC domain was conducted by applying MARCOIL version 1.0 with emission probabilities 9FAM, and the domain was predicted a threshold of 1% to 99%. Subcellular localization of putative NBS genes was analyzed using Target P1.1. The nuclear localization signals (NLSs) of NBS genes were predicted using the NLStradamus program under a two-state HMM static model using posterior and Viterbi methods with 0.6 cut-off value [20].

Structure pattern of Predicted NBS Genes and their location on karyotype

Structure pattern of predicted NBS genes was found by Gene Structure Display Server (GSDS 2.0) (http://gsds.cbi.pku.edu.cn) [21]. Based on the coding regions' architecture, NBS genes were selected for primer designing to amplify the RGAs in *T. alexandrinum*. However, Ensembl Plants (http://plants.ensembl.org) was used to visualize the identified

NBS genes' location on their respective chromosomes.

Selection of *R* genes based on NBS-LRR domain and evolutionary distance

One hundred and thirteen (113) *R* genes of *M. sativa* and *M. truncatula* retrieved from the database (http://www.prgdb.crg.eu) were selected based on the presence of NBS-LRR domain searched by using protein domains finder database, PROSITE (https://prosite.expasy.org) and LRR finder database (http://www.lrrfinder.com/lrrfinder.php). The *R* genes selected based on the NBS-LRR domain's presence were then analyzed to find their genetic distance. Among these, diverse *R* genes were selected for primer designing to identify the RGAs from the genome of *T. alexandrinum*. The cladogram was generated using NBS-LRR domain-containing amino acid sequences of *R* genes with a neighbor-joining approach with the p-distance model. The amino acid sequences were aligned through a multiple sequence alignment tool, ClustalW. The dendrogram was constructed using the MEGA program version 7 [22].

Phylogenetic relationship of NBS genes from *T. pratense*, *M. sativa*, and *M. truncatula*

A phylogenetic tree was constructed to find the phylogenetic relationship of the NBS gene from *T. pratense*, with its closely related species, *M. sativa* and *M. truncatula*. The phylogram was constructed by aligning the protein sequences of selected *R* genes with the NBS-LRR domain of *M. sativa* and *M. truncatula* (reference protein sequences) and protein sequences of putative NBS genes from the *T. pratense* genome. The aligned data matrix was subjected to ModelTest analysis in the MEGA 7 package to find the best evolutionary model with the lowest BIC (Bayesian Information Criterion) value for phylogenetic analysis. The Maximum Likelihood method was used to construct a phylogenetic tree.

Primer designing

Primers were designed through the PrimerQuest Tool and were analyzed using the OligoAnalyzer Tool, provided by Integrated DNA Technologies, Inc. [US]. Primer sequences with approximate amplicon sizes are given in Table S1.

PCR analysis was performed using eight primer pairs designed from the conserved regions of the NBS-LRR domain of *R* genes of *M. sativa* and *M. truncatula* and six primer pairs designed from the conserved regions of putative NBS-LRR genes from the *T. pratense* genome. These primer pairs were used to amplify genomic fragments of NBS-LRR from the *T. alexandrinum* cultivar. DNA was extracted using a DNA extraction kit GeneJET Plant Genomic DNA Purification kit (Thermo Scientific, USA). PCR profile for each primer pair was optimized for the amplification of the product of predicted approximate size. The PCR conditions were standardized to get the required size amplicons. The thermal profile for the amplification of required size DNA fragments was an initial denaturation at 96 °C for 5 min, followed by a loop of 40 cycles comprised of denaturation at 95 °C for 55 s, annealing at 52 ± 3 °C for 30 s, extension at 72 °C for 30 s, and a final extension at 72 °C for 7 min.

Table 1 List of NBS genes on *T. pratense* with Nuclear Localization Signals (NLS).

NBS type	GenBank ID	Nuclear Localization Signal
TIR-NBS-LRR	PNX97834	81–KLKKGKKL–88
	PNX72983	39–KEPEKKKEKPKKEGEKK–55
	PNX75356	39–KEPEKKKEKPKKEGEKK–55
	PNY06688	990–KKSK–993
	PNX76034	122–KHEKRFGKKSEKVK–135
CC-NBS-LRR	PNY07701	93–TRFHPKKILARRNIGKR–109
NBS-LRR	PNY05956	1188–RDELRKRRLR–1197
	PNY05838	2499–RGKEWRK–2505
TIR-NBS	PNY18143	860–RKKSKERGKE–869

RESULTS AND DISCUSSION

The class of NBS-LRR containing proteins consists of four major domains: N-terminal domain, NBS domain, LRR domain, and variable carboxy-terminal domains [23]. The NBS-LRR encoding genes are categorized based on the domain present at the N-terminal of NBS-LRR proteins [24]. Three domains have been found at the N-terminal of NBS-LRR proteins, and they are TIR (Toll/Interleukin-1 receptor), CC (Coiled-coil), and RPW8 (Resistance to powdery mildew8). Based on the presence of these domains, NBS-LRR genes are classified as TNL (TIR-NBS-LRR), CNL (CC-NBS-LRR), and RNL (RPW8-NBS-LRR). However, the NBS domain consists of various conserved motifs including P-loop, GLPL domain, Kinase-2 domain, and kinase-3a domain [13]. The comparison between nucleotide and deduced amino acid sequences of different NBS genes had shown high-level sequence conservation. A diverse array of *R* genes in plants, that had been cloned, showed to have DNA sequence coding for an NBS domain followed by the LRR domain and have been proved to be members of the NBS-LRR superfamily [25, 26]. Therefore, in this study, the genome-wide

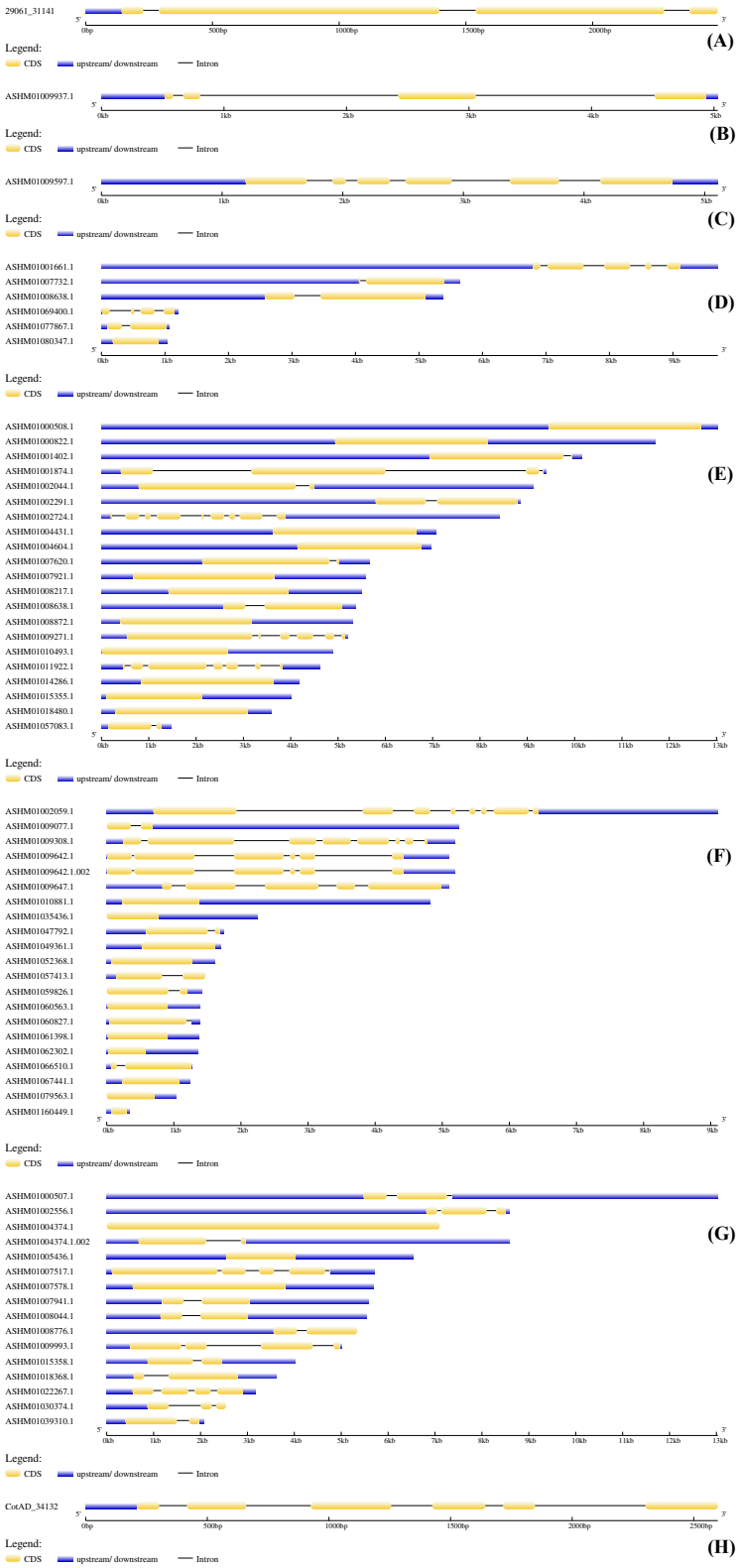

Fig. 1 Gene structures (Exon-intron architecture) of the coding sequence of NBS genes in *T. pratense*: (A) NBS-LRR (NL) genes, (B) RPW8-NBS (RN) genes, (C) RPW8-NBS-LRR (RNL) genes, (D) CC-NBS (CN) genes, (E) CC-NBS-LRR (CNL) genes, (F) NBS (N) genes, (G) TIR-NBS (TN) genes, and (H) TIR-NBS-LRR (TNL) genes.

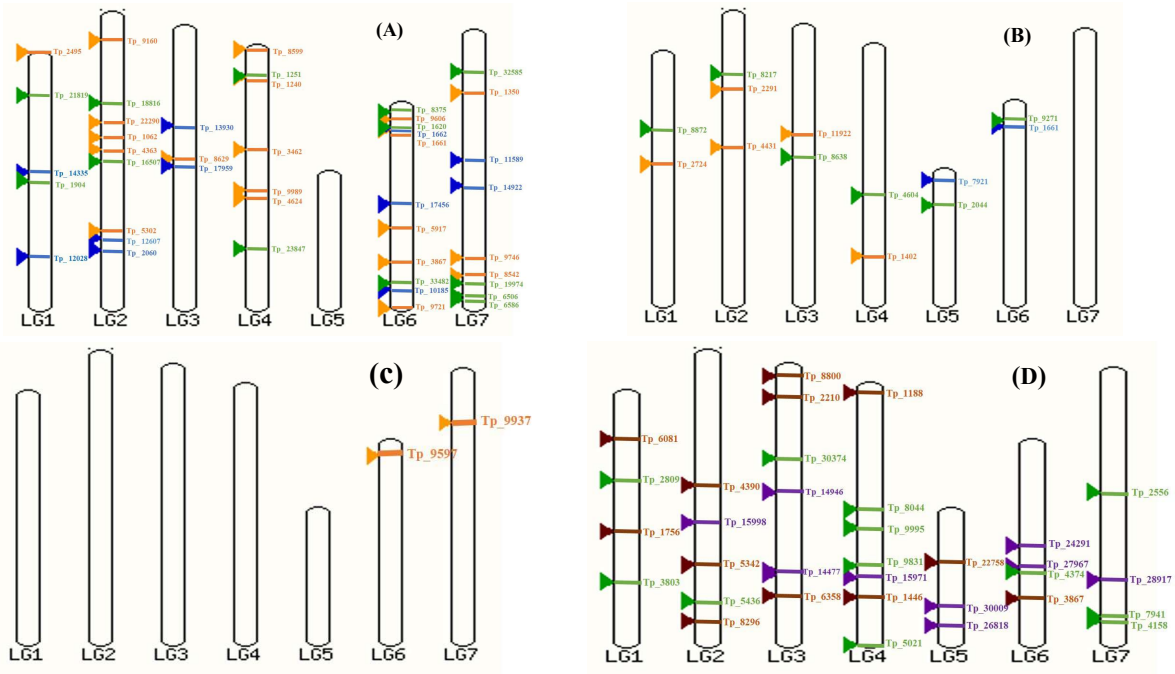

Fig. 2 Chromosomal location of NBS genes: Location of identified NBS genes of different NBS types on the karyotype of *T. pretense*: LG1 = Linkage group1, LG2 = Linkage group2, LG3 = Linkage group3, LG4 = Linkage group4, LG5 = Linkage group5, LG6 = Linkage group6; and LG7 = Linkage group7.

thrash out of *T. pratense* was carried out to identify the NBS-LRR class genes. The InterProScan and PROSITE ID annotation-based analysis of protein-encoding genes of *T. pratense* showed 251 protein sequences were established to have NBS domain (Table S2). Based on the presence of domains, NBS encoding genes in *T. pratense* were categorized into eight groups: CNL (CC-NBS-LRR), CN (CC-NBS), TNL (TIR-NBS-LRR), TN (TIR-NBS), RNL (RPW8-NBS-LRR), RN (RPW8-NBS), NL (NBS-LRR), and N (NBS). Among these, 23 belonged to the CNL group (further validated by MARCOIL server), 74 to the TNL group, 1 to the RNL group, 16 to the TN group, 6 to the CN group, 1 to the RN group, as well as 109 and 21 genes were shown to possess neither CC, TIR nor RPW8 domains, thus grouped as NL (NBS-LRR) and N groups, respectively. Three key signature motifs, P-loop, Kinase-2, and GLPL, of the NBS domain of disease-resistant proteins, were present in all NBS domain-containing proteins. The NBS domain is the highly conserved domain compared with other protein domains encoding by *R* genes and, hence, is used for primer designing to amplify disease resistance-related genes [27].

The localization of identified NBS proteins was predicted through analysis using TargetP software. This analysis predicted the localization of 16 NBS proteins to the mitochondria having mTP signal peptide, 9 to the chloroplast having cTP signal peptide, 44 to the secretory pathway with SP signal peptide, and 182 to subcellular locations other than chloroplast or mitochondria (Table S3). Five TNLs, one CNL, two NLs, and one TN were predicted to have a putative nuclear localization signal (NLS) using the NLStradamus bioinformatics tool (Table 1).

The structure pattern of predicted NBS genes using Gene Structure Display Server revealed the number of exons for CNLs, CNs, TNs, Ns, RNLs, and RNs ranged from 1 to 9, 2 to 4, 1 to 5, 1 to 5, 6, and 4, respectively. However, sequences of the NBS-LRR (NL) group were displayed as single gene architecture having UTRs with four CD regions. Similarly, the TIR-NBS-LRR (TNL) group sequences were also displayed as a single gene pattern comprising 5′UTR, six exons, five introns, and 3′UTR (Fig. 1). On average, CNLs, CNs, TNs, and Ns had 2.6, 2.5, 2.5, and 1.2 exons per gene, respectively. The complete detail of the number of exons and the number of introns against each entry of all NBS types is presented in Table S3. By considering the gene structure display of all 8 groups of NBS genes from *T. pratense*, the NBS-LRR (NL)

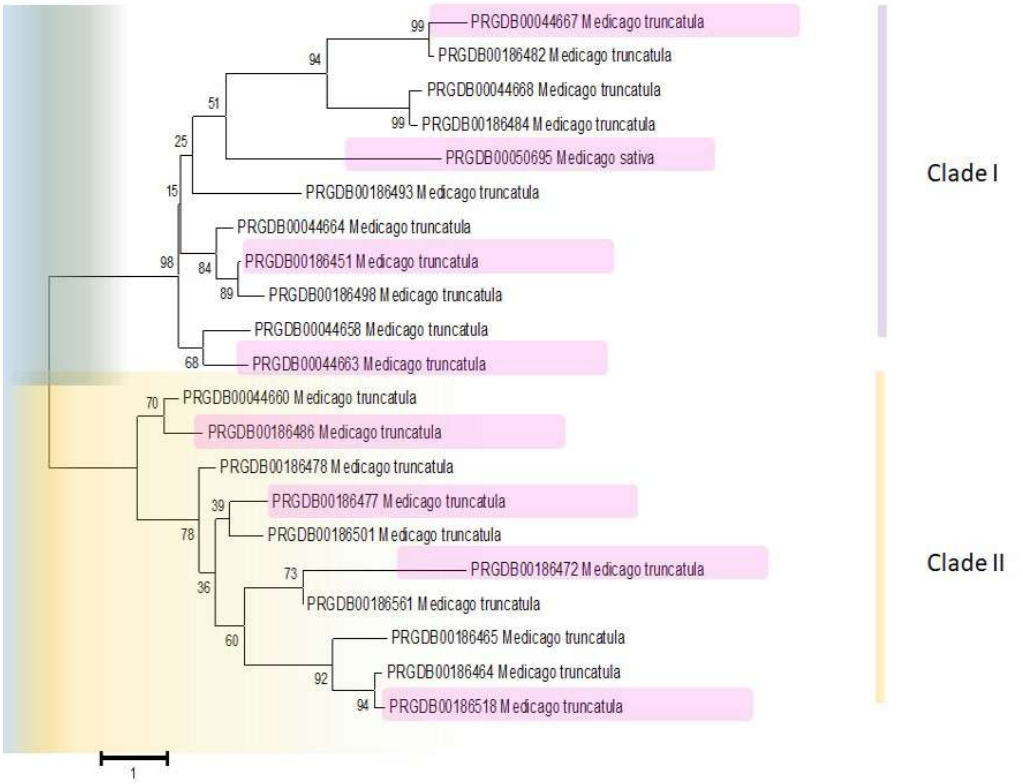

Fig. 3 Cladogram of *R* genes of *M. sativa* and *M. truncatula* possessing NBS-LRR domain generated through the Neighbor-joining method using the p-distance model with 1000 bootstrap replication (MEGA7 software package) showing the genetic distance among them.

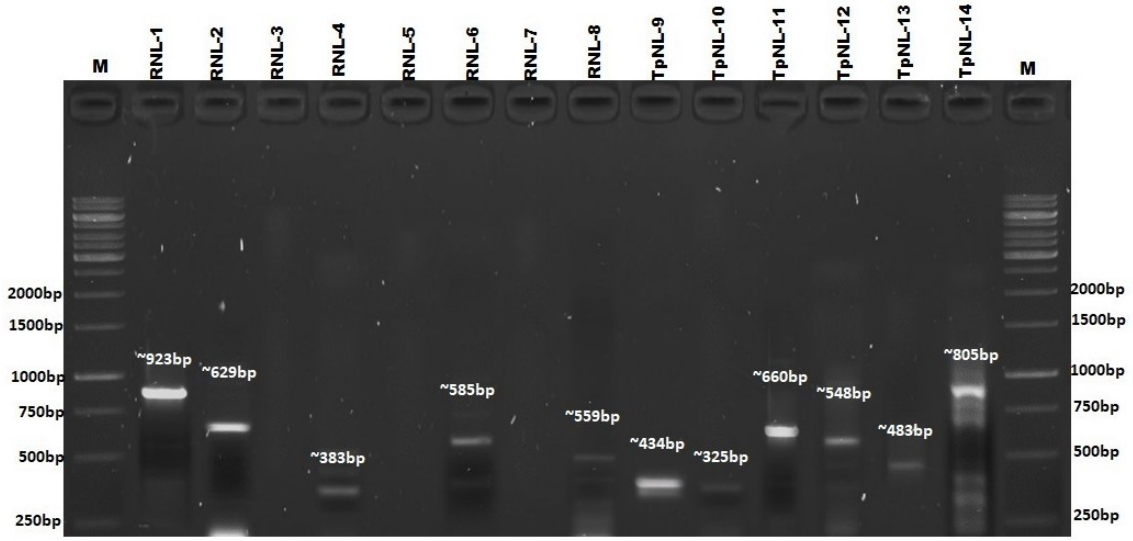

Fig. 4 PCR analysis of *T. alexandrinum* showing cross-species amplification using primer pairs based on NBS-LRR encoding genetic regions of *T. pratense*.

group's conserved sequences were used for primer designing to recover RGAs from *T. alexandrinum*. Besides the structure pattern of all NBS types, the location of each GenBank entry of NBS types: TIR-NBS-LRR (TNL), CC-NBS-LRR (CNL), RWP8-NBS-LRR (RNL), and NBS-LRR (NL), was visualized on karyotype (Fig. 2) through Ensembl Plants portal providing gene IDs given in Table S2.

Out of 113 *R* genes of *M. sativa* and *M. truncatula*, eight (8) *R* genes were not of NBS-LRR class, though 30 had NBS domain, 54 had only LRR domain; however, 21 R genes revealed the presence of NBS-LRR domain in their sequences. Among these 21 *R* genes, 8 were selected based on the evolutionary distance for primer designing to identify *R* genes or RGAs at the genomic level in *T. alexandrinum*. A cladogram of amino acid sequences of regions containing the NBS-LRR domain from selected *R* genes was developed (Fig. 3) through the Neighbor-joining method using the MEGA 7 software package. The evolutionary distances were computed using the JTT matrix-based method [28] and are in the units of the number of amino acid substitutions per site. The rate variation among sites was modeled with a gamma distribution.

A maximum likelihood method based phylogram was constructed for exploring the phylogenetic relationship amongst NBS genes encoding protein sequences of *T. pratense* and *R* genes of *M. truncatula* and *M. sativa* encoding protein sequences having NBS-LRR domain. In maximum likelihood-based phylogenetic analysis, the Jones-Taylor-Thornton (JTT) evolutionary model was selected as the best-fit amino acid substitution model based on the lowest Bayesian information criterion (BIC) value. This phylogram showed the highly close phylogenetic relationship between NBS-LRR regions of *T. pratense*, *M. truncatula*, and *M. sativa*, (**??**). It showed the highly conserved nature of the NBS-LRR class of the *R* gene. Hence, we used the conserved NBS-LRR genetic regions for primer designing to recover and amplify the RGAs from the *T. alexandrinum* genome. It will help to identify the *R* genes in this plant against its phytopathogens. Rachana et al [29] used the conserved motifs of the NBS-LRR domain of resistance genes of date palm for primer designing to recover RGAs from coconut because the genome sequence of coconut was not available in public databases. Here, they used the sequence of different genera for the identification of RGAs in other genera. Various scientists used conserved domains of NBS-LRR genes to recover RGAs or candidate resistance genes from various crop plants, such as citrus, coffee, grape wine, chicory [30, 31], chickpea [32], soybean [33, 34], apple [35], wheat [36], and sorghum [37]. Therefore, cross-species amplification of *in silico* identified NBS-LRR genes based primers was achieved in *T. alexandrinum*. The DNA of a cultivar of *T. alexandrinum*, "Anmol", was subjected to PCR analysis using the primer pairs given in Table S1. It was performed to screen the primers to identify and recover resistance gene sequences in berseem clover in future studies against stem and crown rot diseases. The PCR conditions were standardized to get the required size amplicons. The PCR analysis gave the amplification using all primer pairs except RNL3-F/R, RNL5-F/R, and RNL7-F/R (Fig. 4), showing that they could be used to identify and recover resistance gene sequences in berseem clover under disease challenge.

CONCLUSION

Our study identified 251 NBS domain encoding R proteins through a genome-wide in-silico mining of the *T. pretense* genome. They were further examined for subcellular localization, chromosomal location of NBS domain encoded *R* genes, gene structure, and phylogenetic analysis. The in silico identified NBS-LRR genes based primers were used to analyze their cross-species amplification in *T. alexandrinum*. These findings would help to identify and recover disease-resistance gene sequences in berseem clover under disease challenge.

Appendix A. Supplementary data

Supplementary data associated with this article can be found at http://dx.doi.org/10.2306/scienceasia1513-1874.2021.054. Table S3 can be accessed from the authors.

Acknowledgements: This work was supported by Dr. Siddra Ijaz Laboratory, "Molecular Biology of Plant Disease Resistance Lab" of CABB, University of Agriculture Faisalabad.

REFERENCES

1. Savary S, Bregaglio S, Willocquet L, Gustafson D, D'Croz DM, Sparks A, Castilla N, Djurle A, et al (2017) Crop health and its global impacts on the components of food security. *Food Sec* **9**, 311–327.
2. Pallas V, García JA (2011) How do plant viruses induce disease? Interactions and interference with host components. *J Gen Virol* **92**, 2691–2705.
3. Jones JD, Dangl JL (2006) The plant immune system. *Nature* **444**, 323–329.

4. Li L, Ren M, He F, Xu R (2019) Mapping the stripe rust resistance gene in Chinese wheat Guinong 775. *ScienceAsia* **45**, 318–323.
5. Hammond-Kosack KE, Parker JE (2003) Deciphering plant-pathogen communication: fresh perspectives for molecular resistance breeding. *Curr Opin Biotechnol* **14**, 177–193.
6. van Ooijen G, van den Burg HA, Cornelissen BJ, Takken FL (2007) Structure and function of resistance proteins in solanaceous plants. *Annu Rev Phytopathol* **45**, 43–72.
7. Michelmore RW, Meyers BC (1998) Clusters of resistance genes in plants evolve by divergent selection and a birth-and-death process. *Genome Res* **8**, 1113–1130.
8. Michelmore R (2000) Genomic approaches to plant disease resistance. *Curr Opin Plant Biol* **3**, 125–131.
9. Kang BC, Yeam I, Jahn MM (2005) Genetics of plant virus resistance. *Annu Rev Phytopathol* **43**, 581–621.
10. Young ND (2000) The genetic architecture of resistance. *Curr Opin Plant Biol* **3**, 285–290.
11. Cannon SB, Zhu H, Baumgarten AM, Spangler R, May G, Cook DR, Young ND (2002) Diversity, distribution, and ancient taxonomic relationships within the TIR and non-TIR NBS-LRR resistance gene subfamilies. *J Mol Evol* **54**, 548–562.
12. Pan Q, Wendel J, Fluhr R (2000) Divergent evolution of plant NBS-LRR resistance gene homologs in dicot and cereal genomes. *J Mol Evol* **50**, 203–213.
13. Baldi PA, Patocchi A, Zini E, Toller C, Velasco R, Komjanc M (2004) Cloning and linkage mapping of resistance gene homologs in apple. *Theor Appl Genet* **109**, 231–239.
14. Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant Medicago truncatula. *Plant Physiol* **146**, 5–21.
15. Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. *Plant Cell* **15**, 809–834.
16. Moroldo M, Paillard S, Marconi R, Fabrice L, Canaguier A, Cruaud C, De Berardinis V, Guichard C, et al (2008) A physical map of the heterozygous grapevine 'Cabernet Sauvignon' allows mapping candidate genes for disease resistance. *BMC Plant Biol* **8**, ID 66.
17. Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ (2009) Genome-wide analysis of Carica papaya reveals a small NBS resistance gene family. *Mol Genet Genomics* **281**, 609–626.
18. Tan X, Meyers BC, Kozik A, West MA, Morgante M, St Clair DA, Bent AF, Michelmore RW (2007) Global expression analysis of nucleotide-binding site-leucine rich repeat-encoding and related genes in *Arabidopsis*. *BMC Plant Biol* **7**, ID 56.
19. Zhou T, Wang Y, Chen JQ, Araki H, Jing Z, Jiang K, Shen J, Tian D (2004) Genome-wide identification of NBS genes in japonica rice reveals significant expansion of divergent non-TIR NBS-LRR genes. *Mol Genet Genomics* **271**, 402–415.
20. Ba AN, Pogoutse A, Provart N, Moses AM (2009) NLStradamus: a simple hidden Markov model for nuclear localization signal prediction. *BMC Bioinformatics* **10**, ID 202.
21. Hu B, Jin J, Guo AY, Zhang H, Luo J, Gao G (2014) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* **31**, 1296–1297.
22. Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* **33**, 1870–1874.
23. McHale L, Tan X, Koehl P, Michelmore RW (2006) Plant NBS-LRR proteins: adaptable guards. *Genome biol* **7**, ID S1.
24. Shao ZQ, Xue JY, Wu P, Zhang YM, Wu Y, Hang YY, Wang B, Chen JQ (2016) Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns. *Plant Physiol* **170**, 2095–2109.
25. Ellis J, Dodds P, Pryor T (2000) Structure, function and evolution of plant disease resistance genes. *Curr Opin Plant Biol* **3**, 278–284.
26. Hulbert SH, Webb CA, Smith SM, Sun Q (2001) Resistance gene complexes: evolution and utilization. *Annu Rev Phytopathol* **39**, 285–312.
27. Panwar P, Jha AK, Pandey PK, Gupta AK, Kumar A (2011) Functional markers based molecular characterization and cloning of resistance gene analogs encoding NBS-LRR disease resistance proteins in finger millet (*Eleusine coracana*). *Mol Biol Rep* **38**, 3427–3436.
28. Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* **8**, 275–282.
29. Rachana KE, Naganeeswaran SA, Fayas TP, Thomas RJ, Rajesh MK (2016) Cloning, characterization, and expression analysis of NBS-LRR-type resistance gene analogues (RGAs) in coconut. *Acta Bot Croat* **75**, 1–10.
30. Deng Z, Huang S, Ling P, Chen C, Yu C, Weber CA, Moore GA, Gmitter Jr FG (2000) Cloning and characterization of NBS-LRR class resistance-gene candidate sequences in citrus. *Theor Appl Genet* **101**, 814–822.
31. Donald TM, Pellerone F, Adam-Blondon AF, Bouquet A, Thomas MR, Dry IB (2002) Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. *Theor Appl Genet* **104**, 610–618.
32. Huettel B, Santra D, Muehlbauer F, Kahl G (2002) Resistance gene analogs of chickpea (*Cicer arietinum*

L.): isolation, genetic mapping, and association with a Fusarium resistance gene cluster. *Theor Appl Genet* **105**, 479–490.

33. Penuela S, Danesh D, Young ND (2002) Targeted isolation, sequence analysis, and physical mapping of nonTIR NBS-LRR genes in soybean. *Theor Appl Genet* **104**, 261–272.
34. He CY, Tian AG, Zhang JS, Zhang ZY, Gai JY, Chen SY (2003) Isolation and characterization of a full-length resistance gene homolog from soybean. *Theor Appl Genet* **106**, 786–793.
35. Lee SY, Seo JS, Rodriguez-Lanetty M, Lee DH (2003) Comparative analysis of superfamilies of NBS-encoding disease resistance gene analogs in cultivated and wild apple species. *Mol Genet Genomics* **269**, 101–108.
36. Lacock L, van Niekerk C, Loots S, du Preez F, Botha AM (2003) Functional and comparative analysis of expressed sequences from *Diuraphis noxia* infested wheat obtained utilizing the conserved nucleotide binding site. *Afr J Biotechnol* **2**, 75–81.
37. Totad AS, Fakrudin B, Kuruvinashetti MS (2005) Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench). *Euphytica* **143**, 179–188.

Appendix A. Supplementary data

Table S1 List of primer pairs designed from NBS-LRR genetic regions of *T. pratense* and *R* genes of *M. truncatula* as well as *M. sativa* to RGAs from *T. alexandrinum*.

Sr#	Primer name	Primer Sequence (5′ to 3′ direction)	Length (base pair)	Amplicon size
1	RNL1-F	ACAACCCTTGCTACGACTTTAT	22	∼923
	RNL1-R	CTATTGGGTGGGAACAAGATGA	22	
2	RNL2-F	GGTGGAATTGGAAAGACAACAC	22	∼629
	RNL2-R	CCCTTCCAATCTATCACCTCTTC	23	
3	RNL3-F	TAAGGGAAGCAAAATTCTTG	20	∼814
	RNL3-R	CAAATTCAACAAAATGATAT	20	
4	RNL4-F	AATCATCACGACCAGGGATAAG	22	∼383
	RNL4-R	CTTAATTGGGATGGACCGAAATG	23	
5	RNL5-F	TCCAAAGCTTGTGAATATTT	20	∼717
	RNL5-R	GCTGGTGCATCCTTACGGTA	20	
6	RNL6-F	GCAGAGACTACCATGACCTAAC	22	∼585
	RNL6-R	TGATTGTGGTGAAGACCCTG	20	
7	RNL7-F	TGCCAAGGGGTTCCACTAGC	20	∼651
	RNL7-R	CGAAGAATATCTTAAATTCA	20	
8	RNL8-F	CTGCCCGTATATTGTCGGTATC	22	∼559
	RNL8-R	GTCGGTATAGCTTCCAAGTCAG	22	
9	TpNL9-F	TCCGAGTCCAGTCACCTAAT	20	∼434
	TpNL9-R	GGTTAAGCTACCCAAACCTCTC	22	
10	TpNL10-F	CATCTTGTTCCCACCTCACTT	21	∼325
	TpNL10-R	TGCTCGCTCTGCTCTTATTG	20	
11	TpNL11-F	GAAACGGTCTCCAAGGGAATAG	22	∼660
	TpNL11-R	CAACAGCAAGAGCCTGGTATAG	22	
12	TpNL12-F	GGGACAAAGGAAGAGGAAGAAG	22	∼548
	TpNL12-R	GACAAGGTTAGGATGGTGAAGTAG	24	
13	TpNL13-F	CTGCAAAGGGATTGGAGTATCT	22	∼483
	TpNL13-R	CAGCTATAACCGGACGCATATTA	23	
14	TpNL14-F	GATGAGATTTGGTGCTGGTATTG	23	∼805
	TpNL14-R	TATGACCGGAGGATTAGCTTTG	22	

Table S2 List of protein encoding genes of *Trifolium pratense* revealed NBS-LRR class of *R* genes based on annotation analysis using InterProScan and PROSITE ID and MARCOIL programs.

Sr#	GenBank protein ID	Domain	Sr#	GenBank protein ID	Domain	Sr#	GenBank protein ID	Domain
1	PNY18145	TIR-NBS-LRR	15	PNY17715	NBS-LRR	29	PNY10502	NBS-LRR
2	PNY18143	TIR-NBS	16	PNY17706	RPW8-NBS-LRR	30	PNY10903	TIR-NBS-LRR
3	PNY18138	NBS-LRR	17	PNY17373	NBS	31	PNY11184	NBS-LRR
4	PNY18082	RPW8-NBS	18	PNY17240	TIR-NBS-LRR	32	PNY05838	NBS-LRR
5	PNY17495	TIR-NBS-LRR	19	PNY08509	TIR-NBS	33	PNY09918	NBS-LRR
6	PNY17967	TIR-NBS-LRR	20	PNY12075	TIR-NBS-LRR	34	PNY06933	NBS-LRR
7	PNY17919	NBS-LRR	21	PNY01589	TIR-NBS-LRR	35	PNY05168	NBS-LRR
8	PNY17876	NBS-LRR	22	PNY05956	NBS-LRR	36	PNY14991	TIR-NBS-LRR
9	PNY17871	NBS-LRR	23	PNY04299	NBS-LRR	37	PNX77738	NBS-LRR
10	PNY17847	NBS-LRR	24	PNY05217	NBS-LRR	38	PNX74726	TIR-NBS-LRR
11	PNY17844	NBS-LRR	25	PNY08911	TIR-NBS-LRR	39	PNX98847	TIR-NBS-LRR
12	PNY17754	NBS	26	PNY09295	TIR-NBS-LRR	40	PNX98800	TIR-NBS-LRR
13	PNY17750	NBS	27	PNY10327	TIR-NBS-LRR	41	PNX98321	NBS-LRR
14	PNY17749	NBS	28	PNY10412	TIR-NBS-LRR	42	PNX72983	TIR-NBS-LRR

Table S2 Continued ...

Sr#	GenBank protein ID	Domain	Sr#	GenBank protein ID	Domain	Sr#	GenBank protein ID	Domain
43	PNX94829	TIR-NBS-LRR	96	PNY17136	NBS-LRR	149	PNX96840	NBS-LRR
44	PNX94857	TIR-NBS-LRR	97	PNX92863	NBS-LRR	150	PNX95922	NBS-LRR
45	PNX95230	NBS-LRR	98	PNY16865	CC-NBS-LRR	151	PNX95206	NBS-LRR
46	PNX75019	TIR-NBS-LRR	99	PNY16084	CC-NBS-LRR	152	PNX93451	NBS-LRR
47	PNX96830	TIR-NBS-LRR	100	PNY15377	CC-NBS-LRR	153	PNX93207	NBS-LRR
48	PNX96861	TIR-NBS-LRR	101	PNY08769	CC-NBS-LRR	154	PNX92872	NBS-LRR
49	PNY11200	TIR-NBS	102	PNY08097	CC-NBS-LRR	155	PNX85192	NBS-LRR
50	PNX85428	TIR-NBS-LRR	103	PNY08096	CC-NBS-LRR	156	PNX83195	NBS-LRR
51	PNX84848	TIR-NBS-LRR	104	PNY07417	CC-NBS-LRR	157	PNX77817	TIR-NBS-LRR
52	PNX84569	NBS-LRR	105	PNX97258	CC-NBS-LRR	158	PNY15206	NBS-LRR
53	PNX83826	NBS-LRR	106	PNX96277	NBS-LRR	159	PNY13999	NBS-LRR
54	PNX83797	TIR-NBS-LRR	107	PNX96227	CC-NBS-LRR	160	PNY16262	NBS-LRR
55	PNX83539	TIR-NBS-LRR	108	PNX93989	NBS-LRR	161	PNX95529	NBS-LRR
56	PNX57043	NBS-LRR	109	PNX93872	CC-NBS-LRR	162	PNX95461	NBS-LRR
57	PNX57104	NBS-LRR	110	PNY05471	CC-NBS-LRR	163	PNY13233	NBS-LRR
58	PNX73969	TIR-NBS-LRR	111	PNY07701	CC-NBS-LRR	164	PNY06993	NBS-LRR
59	PNX82572	NBS-LRR	112	PNY06519	NBS-LRR	165	PNX85687	NBS
60	PNX81696	NBS-LRR	113	PNX58122	CC-NBS	166	PNX85753	NBS-LRR
61	PNX81644	TIR-NBS-LRR	114	PNX54860	CC-NBS	167	PNX89958	NBS
62	PNY15416	NBS-LRR	115	PNX54307	NBS	168	PNX92002	NBS-LRR
63	PNY14099	NBS-LRR	116	PNY03163	NBS-LRR	169	PNX92329	CC-NBS-LRR
64	PNY14049	NBS-LRR	117	PNY02529	NBS-LRR	170	PNY15513	CC-NBS
65	PNY12462	NBS-LRR	118	PNY05471	CC-NBS-LRR	171	PNY16719	NBS-LRR
66	PNY07728	NBS	119	PNY04848	CC-NBS-LRR	172	PNY16721	NBS-LRR
67	PNY07472	NBS-LRR	120	PNX97069	TIR-NBS-LRR	173	PNY16530	NBS-LRR
68	PNY07284	TIR-NBS-LRR	121	PNX97334	TIR-NBS-LRR	174	PNY16991	NBS-LRR
69	PNY06991	NBS-LRR	122	PNX97809	TIR-NBS-LRR	175	PNY16563	NBS-LRR
70	PNY06326	NBS-LRR	123	PNX97834	TIR-NBS-LRR	176	PNY16572	CC-NBS
71	PNY00116	NBS-LRR	124	PNX94561	NBS-LRR	177	PNY16573	CC-NBS-LRR
72	PNY15963	TIR-NBS-LRR	125	PNX71195	TIR-NBS-LRR	178	PNY12514	TIR-NBS-LRR
73	PNY15875	TIR-NBS	126	PNX93012	TIR-NBS-LRR	179	PNY13450	TIR-NBS-LRR
74	PNY15751	TIR-NBS	127	PNX92725	TIR-NBS-LRR	180	PNY13802	TIR-NBS-LRR
75	PNY15325	TIR-NBS	128	PNX92432	TIR-NBS-LRR	181	PNX92407	TIR-NBS-LRR
76	PNY15249	TIR-NBS	129	PNX92402	TIR-NBS-LRR	182	PNY06688	TIR-NBS-LRR
77	PNY04648	NBS-LRR	130	PNX91618	NBS-LRR	183	PNY07227	TIR-NBS-LRR
78	PNY03145	TIR-NBS	131	PNX91305	NBS-LRR	184	PNY10503	TIR-NBS-LRR
79	PNY02957	NBS-LRR	132	PNX78094	NBS-LRR	185	PNY11224	TIR-NBS-LRR
80	PNY02794	NBS-LRR	133	PNX75625	NBS-LRR	186	PNY05216	TIR-NBS-LRR
81	PNX81137	TIR-NBS	134	PNX76034	TIR-NBS-LRR	187	PNY14743	TIR-NBS-LRR
82	PNX79038	NBS-LRR	135	PNX71791	NBS-LRR	188	PNY11984	TIR-NBS-LRR
83	PNX76273	TIR-NBS	136	PNX71853	NBS-LRR	189	PNY08414	NBS-LRR
84	PNX90271	NBS	137	PNX63972	NBS	190	PNY06204	TIR-NBS-LRR
85	PNY17252	TIR-NBS-LRR	138	PNY12647	TIR-NBS	191	PNY06308	NBS-LRR
86	PNY16749	TIR-NBS	139	PNY11201	TIR-NBS	192	PNY06686	TIR-NBS-LRR
87	PNY17293	NBS-LRR	140	PNX99969	TIR-NBS	193	PNY07957	TIR-NBS-LRR
88	PNY17328	NBS-LRR	141	PNY00065	CC-NBS-LRR	194	PNY04129	TIR-NBS-LRR
89	PNY17329	CC-NB-LRR	142	PNY16463	NBS-LRR	195	PNY04749	TIR-NBS-LRR
90	PNY16780	TIR-NBS-LRR	143	PNY15728	CC-NBS-LRR	196	PNY03546	TIR-NBS-LRR
91	PNY17192	NBS-LRR	144	PNY11560	NBS-LRR	197	PNY04073	TIR-NBS-LRR
92	PNY17101	NBS	145	PNY11532	CC-NBS-LRR	198	PNY04075	TIR-NBS-LRR
93	PNY16929	NBS-LRR	146	PNX58904	NBS	199	PNY04128	TIR-NBS-LRR
94	PNX88689	NBS	147	PNX99616	NBS-LRR	200	PNY01932	IR-NBS-LRR
95	PNX92059	NBS-LRR	148	PNX99165	NBS-LRR	201	PNY16167	TIR-NBS-LRR

Table S2 Continued ...

Sr#	GenBank protein ID	Domain	Sr#	GenBank protein ID	Domain	Sr#	GenBank protein ID	Domain
202	PNY01326	TIR-NBS-LRR	219	PNX63013	NBS-LRR	236	PNX96411	TIR-NBS-LRR
203	PNY00455	TIR-NBS-LRR	220	PNY11278	CC-NBS-LRR	237	PNX95888	NBS-LRR
204	PNY05231	NBS-LRR	221	PNY07064	CC-NBS	238	PNY12037	TIR-NBS-LRR
205	PNX90059	NBS-LRR	222	PNY06607	CC-NBS-LRR	239	PNY11115	NBS-LRR
206	PNX75356	TIR-NBS-LRR	223	PNX92758	NBS	240	PNY11027	NBS-LRR
207	PNX72165	TIR-NBS-LRR	224	PNX91520	NBS	241	PNY10277	TIR-NBS-LRR
208	PNX88855	NBS-LRR	225	PNX89769	NBS	242	PNY09617	NBS-LRR
209	PNX88702	TIR-NBS-LRR	226	PNX88574	CC-NBS-LRR	243	PNY07730	NBS-LRR
210	PNX88691	NBS-LRR	227	PNX86873	NBS	244	PNY07068	NBS-LRR
211	PNX88232	NBS-LRR	228	PNX85024	NBS	245	PNY04845	TIR-NBS
212	PNX54440	TIR-NBS-LRR	229	PNX79175	NBS	246	PNX95556	NBS-LRR
213	PNX87964	NBS-LRR	230	PNX59264	CC-NBS	247	PNX94563	NBS-LRR
214	PNX87325	NBS-LRR	231	PNY01330	NBS-LRR	248	PNX93523	NBS-LRR
215	PNX87250	NBS-LRR	232	PNY00341	NBS-LRR	249	PNX89451	NBS
216	PNX68675	NBS-LRR	233	PNX97851	TIR-NBS-LRR	250	PNX89681	NBS
217	PNX70277	NBS-LRR	234	PNX97261	TIR-NBS	251	PNX77437	NBS-LRR
218	PNX62031	NBS-LRR	235	PNX96655	NBS-LRR			

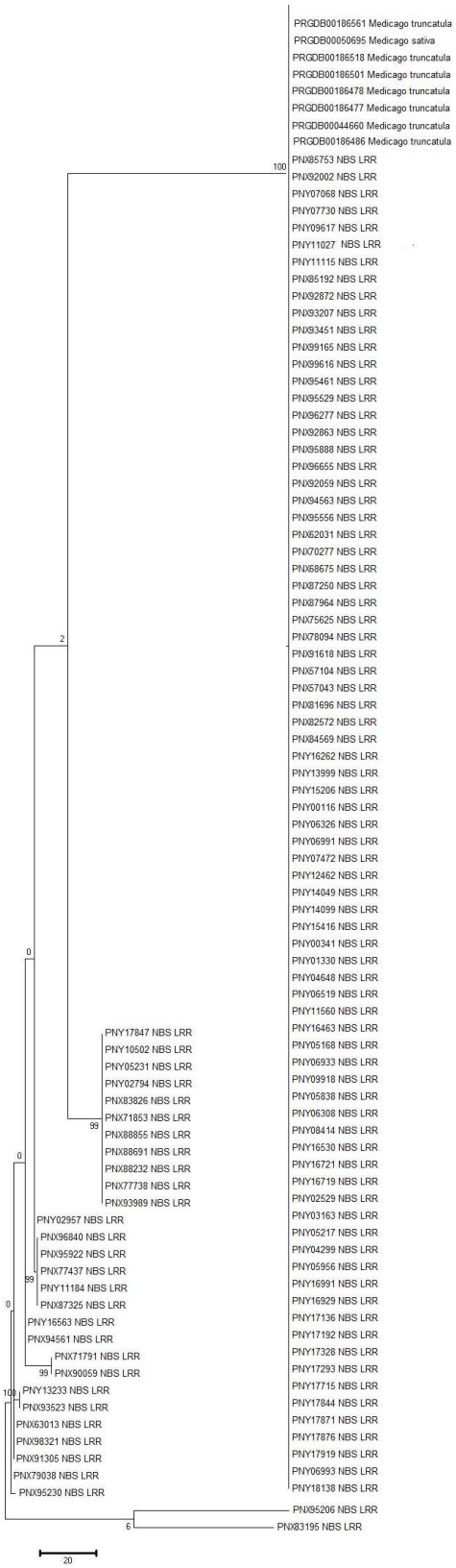

Fig. S1 A phylogram generated using Maximum Likelihood (ML) method based on Jones-Taylor-Thornton (JTT) evolutionary model with 1000 bootstrap replications showing very close relationship among the NBS-LRR genes of Trifolium pratense, *Medicago sativa* and *Medicago truncatula*.