Pakistan Journal of Medical & Cardiological Review
https://pakjmcr.com/index.php/1/about
Online ISSN 3007-2387
Print ISSN 3007-2379

Genome-Wide Identification and Comparative Evolutionary Analysis of NLR Genes in Cajanus cajan

Laraib Safdar
Center of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad, Pakistan
Sahar u Nisa Pechuho
Department of Physiology, University of Sindh, Jamshoro, Pakistan
Iqra Noor
Department of Biochemistry, Government College University, Faisalabad, Pakistan
Rida Younas
Institute of Soil and Environmental Sciences, University of Agriculture Faisalabad, Pakistan
Faran Durrani
Department of Plant Biodiversity, University of Science and Technology, Bannu, Khyber Pakhtunkhwa, Pakistan
Sadia Iqbal
Department of Zoology, Lahore College for Women University, Pakistan

Author Details

Keywords: Cajanus Cajan, Nlr Genes, Evolutionary Analysis, Disease Resistance, Phylogenetics, Gene Duplication, Plant Immunity.

Received on 01 Jan 2026
Accepted on 01 Feb 2026
Published on 11 Feb 2026

Corresponding E-mail & Author*:

Laraib Safdar
Center of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad, Pakistan

Abstract

Cajanus cajan (L.) commonly known as pigeon pea is a vital leguminous crop renowned for its nutritional and agronomic value. However, its productivity is severely constrained by biotic stresses, particularly the pod borer Helicoverpa armigera. Understanding the genetic basis of host resistance is essential for developing resilient cultivars. In this study, a comprehensive genome-wide exploration of nucleotide-binding site leucine-rich repeat (NLR) genes was conducted in C. cajan to elucidate their structural diversity and evolutionary patterns. Using the NLRtracker pipeline integrated with the RefPlantNLR database, 205 NLR homologs were identified and classified into four subclasses: CC-NLR, TIR-NLR, CCG10-NLR and CCR-NLR. Among these, CC-NLR (84 genes) and TIR-NLR (88 genes) were predominant indicating their major role in immune signaling. Multiple sequence alignment revealed conserved FLY motifs within the NBARC domain, suggesting evolutionary conservation across NLR families. Phylogenetic and duplication analyses highlighted dynamic gene birth–death events and lineage-specific expansions, reflecting adaptive diversification under selective pressure. These findings enhance the understanding of NLR evolution in C. cajan and provide a valuable foundation for molecular breeding aimed at improving disease resistance and crop productivity.

Introduction

Plants have evolved sophisticated immune systems to recognize and defend against a broad spectrum of pathogens. A key component of this defense is mediated by the nucleotide-binding site leucine-rich repeat (NLR) gene family, which encodes

intracellular immune receptors responsible for detecting pathogen effectors and activating defense responses (Marone et al., 2013). Structurally, NLR proteins consist of a conserved nucleotide-binding (NBARC) domain, variable N-terminal domains such as Toll/interleukin-1 receptor (TIR) or coiled-coil (CC) motifs and a C-terminal leucine-rich repeat (LRR) region that determines pathogen recognition specificity. These multidomain architectures enable plants to trigger rapid and precise immune signaling pathways (Steele, Hughes and Banfield, 2019).

Cajanus cajan (pigeon pea) is a major grain legume cultivated widely across tropical and subtropical regions. It serves as an important source of dietary protein contributes to soil nitrogen fixation and supports agricultural sustainability (Kushwaha and Mehta, 2023). Despite its economic and nutritional importance, pigeon pea productivity is frequently reduced by biotic stresses such as Fusarium wilt, sterility mosaic disease and phytophthora blight (Keerthi et al., 2022). Traditional breeding approaches for disease resistance have been limited by narrow genetic diversity and a lack of molecular understanding of resistance gene evolution (Singer et al., 2021). Therefore, identifying and characterizing NLR genes in C. cajan is essential for developing durable, broad-spectrum resistance through molecular breeding and genomic-assisted selection.

Recent advances in whole-genome sequencing and bioinformatics tools have enabled systematic genome-wide identification and evolutionary analysis of NLR genes in many crop species. Comparative genomics and phylogenetic analyses provide valuable insights into how gene duplication, diversification and selection pressures shape the NLR repertoire across lineages (Santos et al., 2022). Understanding these evolutionary trajectories can uncover functionally significant variants and contribute to improved disease management strategies.

The present study aims to explore the NLR gene family in Cajanus cajan through genome-wide identification, domain-based characterization and phylogenetic analysis. Furthermore, comparative evolutionary analyses with related legume species were conducted to trace the diversification and adaptation patterns of NLR genes. These findings are expected to enhance the understanding of plant immune gene evolution and provide genomic resources for resistance breeding in pigeon pea.

Materials and Methods

Identification and Annotation of NLR Genes

NLR genes were identified in the Cajanus cajan proteome using NLRtracker, a pipeline based on the RefPlantNLR dataset (Kourelis et al., 2021). This tool integrates InterProScan (v5.51-85.0) for domain prediction and FIMO from the MEME suite (v5.1.1) to detect predefined NLR motifs. Core domains including NBARC, LRR, TIR, RPW8, CC and integrated domains (e.g., late blight R1) were recognized and annotated. Functional annotations were refined using the refplantNLR R package (Kourelis et al., 2021).

For comparative phylogenetic analyses, NBARC domain sequences were retrieved directly from NLRtracker outputs ensuring accurate domain architecture reconstruction.

RefPlantNLR Database Utilization

The RefPlantNLR database containing 481 experimentally validated NLR sequences from 31 angiosperm genera (Kourelis et al., 2021) was used as a reference for annotation and classification. This dataset includes diverse NLR subclasses—TIR-NLR, CC-NLR, CCR-NLR and CCG10-NLR—representing major clades such as Arabidopsis, Solanaceae and cereals. RefPlantNLR entries were functionally annotated using InterProScan, LRR predictor and HMMAU allowing identification of canonical and non-canonical integrated domains. Phylogenetic clustering confirmed alignment with previously reported NLR evolutionary groups.

Sequence Extraction and Processing

Protein sequences of C. cajan were retrieved from the GigaScience Database (release 2020-06-09; file size: 11.63 MB). FASTA headers were cleaned and validated before analysis (Xiao et al., 2019). Outputs were verified using standard Linux utilities (cat wget-log). Redundant sequences were clustered using CD-HIT at 90% identity and representative sequences were selected for each cluster using a custom R script.

Multiple Sequence Alignment and Phylogenetic Analysis

Amino acid sequences of NLR and NBARC domains were aligned using MUSCLE integrated within AliView, a lightweight Java-based alignment editor supporting multiple formats (Edgar, 2004). Gaps were manually curated and alignments were exported in MUSCLE format for further analysis. Maximum-likelihood phylogenetic trees were constructed using IQ-TREE (v2.0) under the JTT+F+R4 substitution model with 1,000 bootstrap replicates. Reference NBARC domains from the PRG database were clustered using UCLUST (50% identity) to serve as seed probes for phylogenetic classification (Prasad, Madhusudanan and Jaganathan, 2015).

Chromosomal Localization and Gene Density Mapping

Chromosomal positions of identified NLR genes were determined using BED tools ("make-windows" and "intersect" functions) with 5 kb bins. Unplaced scaffolds were excluded to ensure mapping accuracy. Gene density maps were visualized using Rideogram, which generated linearized chromosome diagrams illustrating NLR distribution across the C. cajan genome. Syntenic relationships and collinearity between NLR loci were analyzed using BLAST and genomic coordinates from the GTF annotation file (Madden, 2013).

Evolutionary and Comparative Genomic Analyses

To evaluate gene duplication and selection pressure, paralogous NLR coding sequences were aligned using ClustalW and corresponding nucleotide alignments were generated using PAL2NAL. The Ka/Ks ratio was computed with the Ka/Ks Calculator (Hung and Weng, 2016) and paralogs with $P > 0.01$ (Fisher's test) were considered under neutral selection.

Orthologous clustering among NLR genes from Cajanus cajan and related legumes (Phaseolus, Glycine, Medicago) was performed using OrthoVenn2 and OrthoFinder under default parameters (E-value = 1e−2). Gene family expansion and contraction analyses were conducted using CAFE (v5.3.0) and the resulting orthogroup data were converted into ultrametric trees using the APE package in R.

Results

Conserved Regions of Cajanus cajan

Alignment of NBARC domain sequences using the MUSCLE algorithm integrated into the AliView tool revealed several conserved motifs across C. cajan NLRs. Among these, the "FLY" motif showed the highest degree of conservation, suggesting its crucial functional role in ATP binding and signal transduction (fig. 1). These conserved regions are characteristic signatures of functional NLRs and support their evolutionary conservation among legume species.

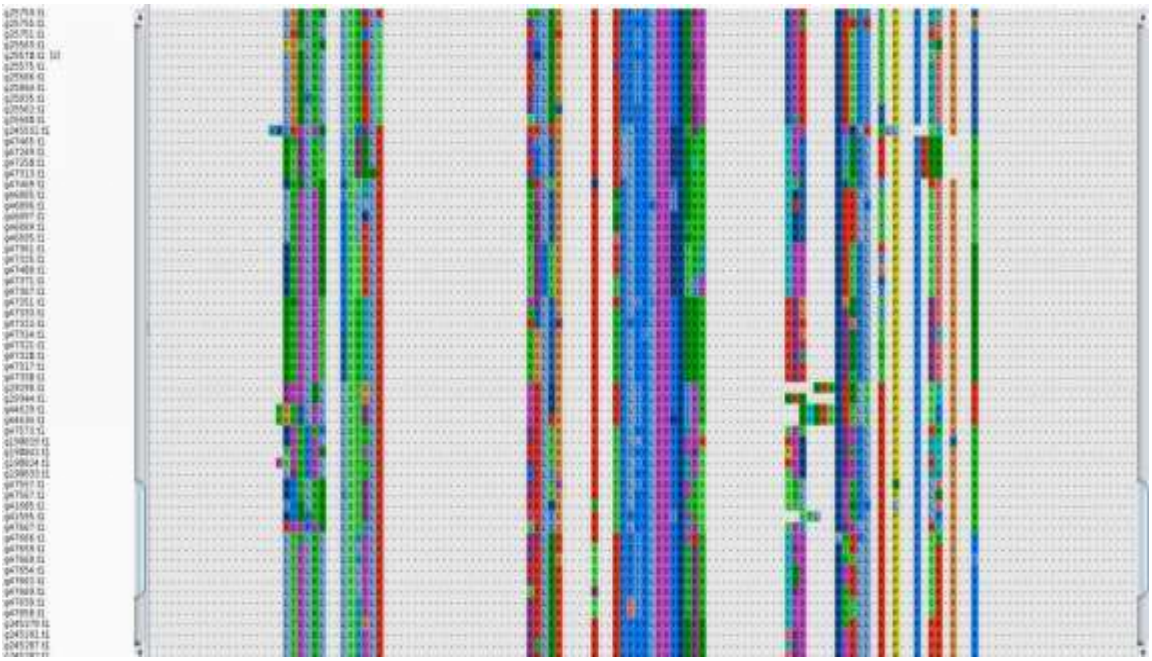

Fig. 1: Identification and characterization of conserved NLR regions in C. cajan

Gene Mining

Genome-wide mining of the C. cajan genome identified approximately 205 NLR homologs distributed across four main classes: CC-NLR (84 genes), TIR-NLR (88 genes), CCG10-NLR (29 genes) and CCR-NLR (the least represented). The predominance of CC and TIR-type NLRs indicates extensive diversification of these classes in the C. cajan genome (fig. 2).

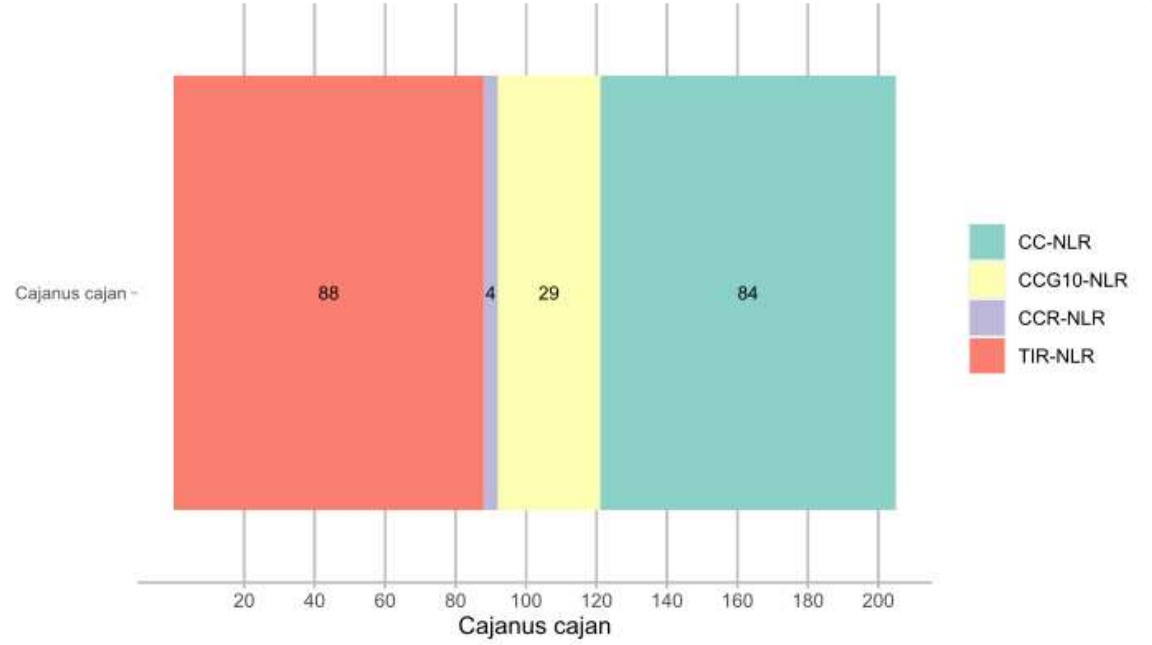

Fig. 2: NLR gene mining in C. cajan

Analysis of NLR length distribution showed that CC-NLRs exhibited the greatest sequence length (>20) followed by CCG10-NLRs, whereas CCR-NLRs were comparatively shorter (fig. 3). Similarly, NBARC domain length analysis showed that CC-NLRs possessed longer NBARC domains (>40) suggesting structural complexity and potential for functional diversity (fig. 4).

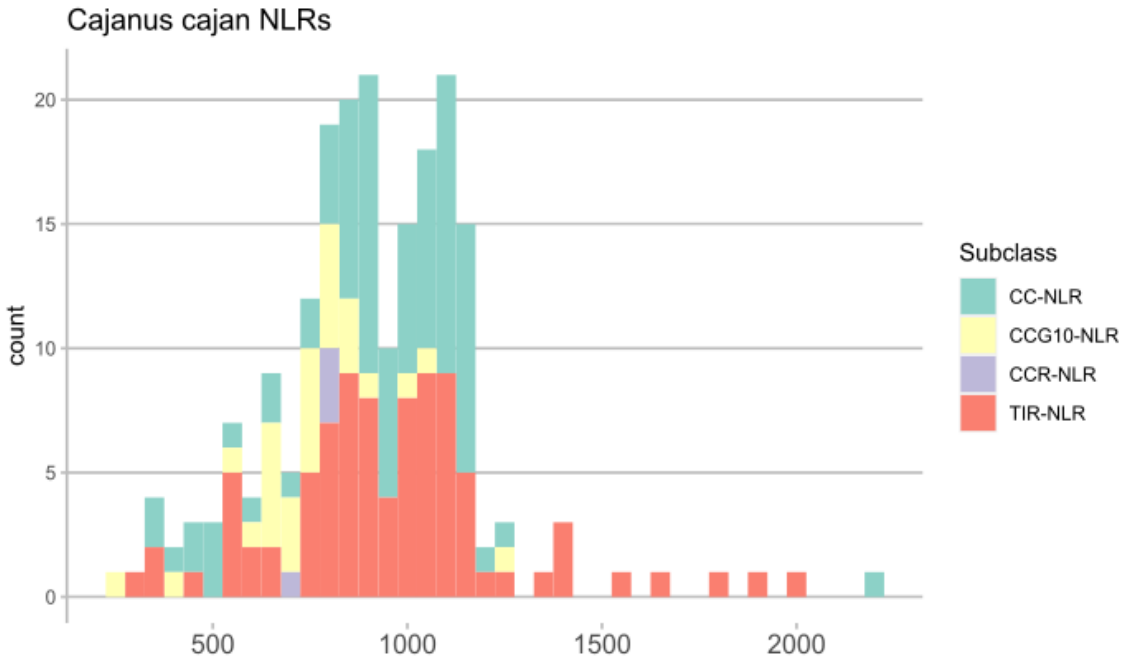

Fig. 3: NLR sequence length of C. cajan

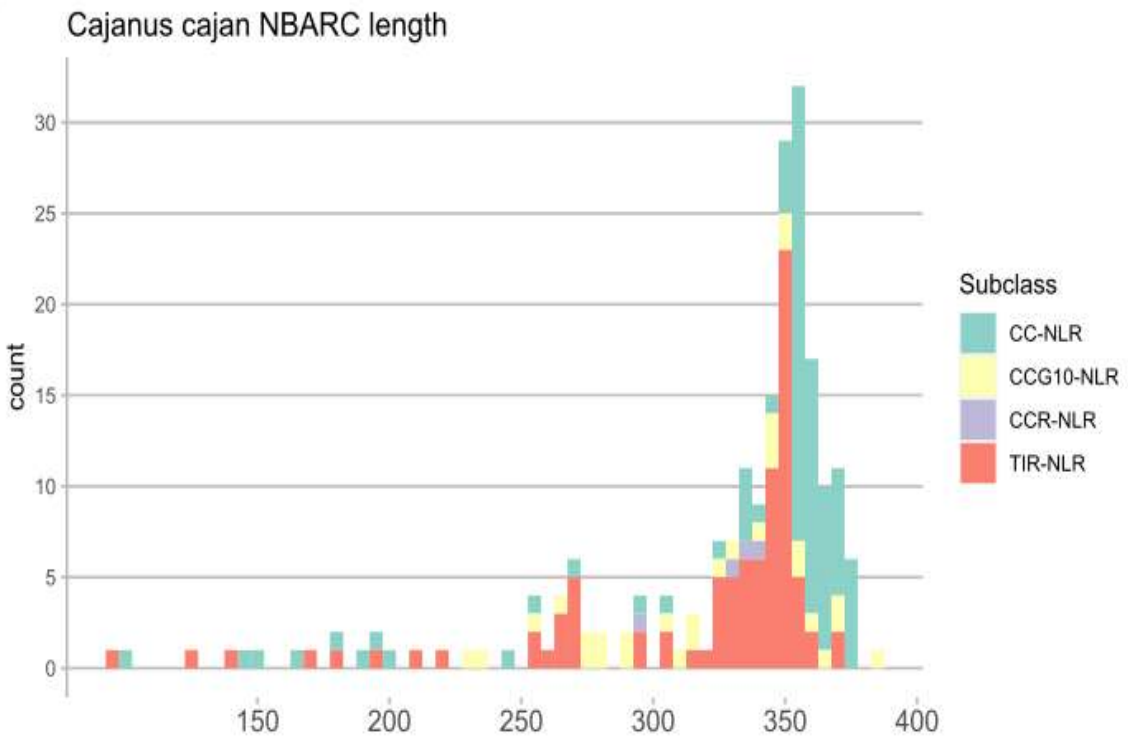

Fig. 4: NBARC sequence length of C. cajan

Domain architecture analysis (fig. 5) indicated that the CNL (CC-NLR) class appeared most frequently followed by TNL (TIR-NLR), consistent with previous findings in other legume genomes.

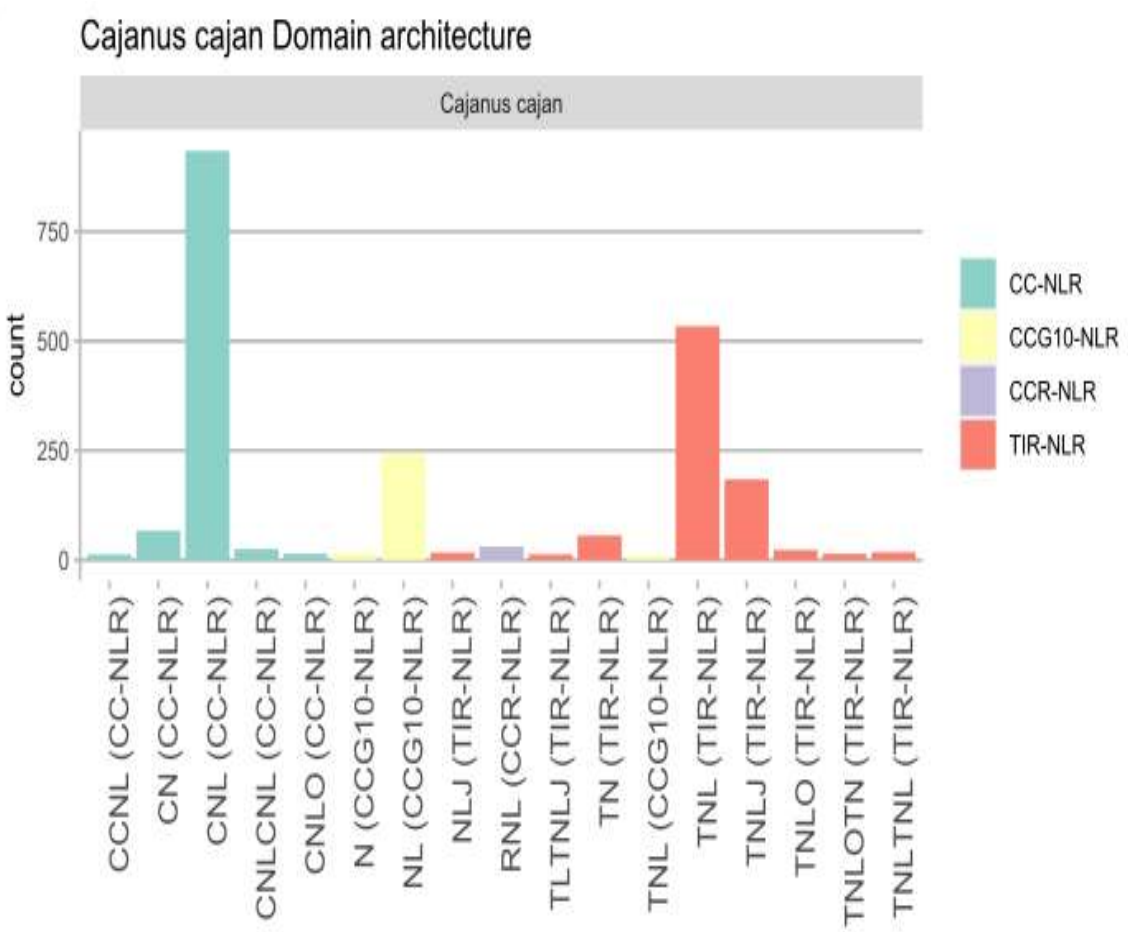

Fig. 5: NLR domain structure in C. cajan

Phylogenetic Distribution

Phylogenetic clustering of conserved NBARC domains using CD-HIT (75% sequence identity) revealed distinct clades within the C. cajan NLR repertoire. The TNL clade exhibited four radiations and remained polyphyletic, whereas CC-NLR, CCR-NLR and CCG10-NLR formed three major monophyletic subclades. CC-NLRs were further divided into four groups; CNL-Un, CNL-G11, CNL-G7 and G4 demonstrating clear lineage-specific expansion.

CCG10-NLRs displayed two prominent subclades, reflecting high diversification within Fabaceae. Interestingly, C. cajan lacked CNL groups G1–G3, G6 and G8, consistent with earlier observations in Cicer species. The predominance of TIR and CC-NLR genes suggests that C. cajan experienced imbalanced gene duplication events following its divergence from a common ancestor (fig. 6).

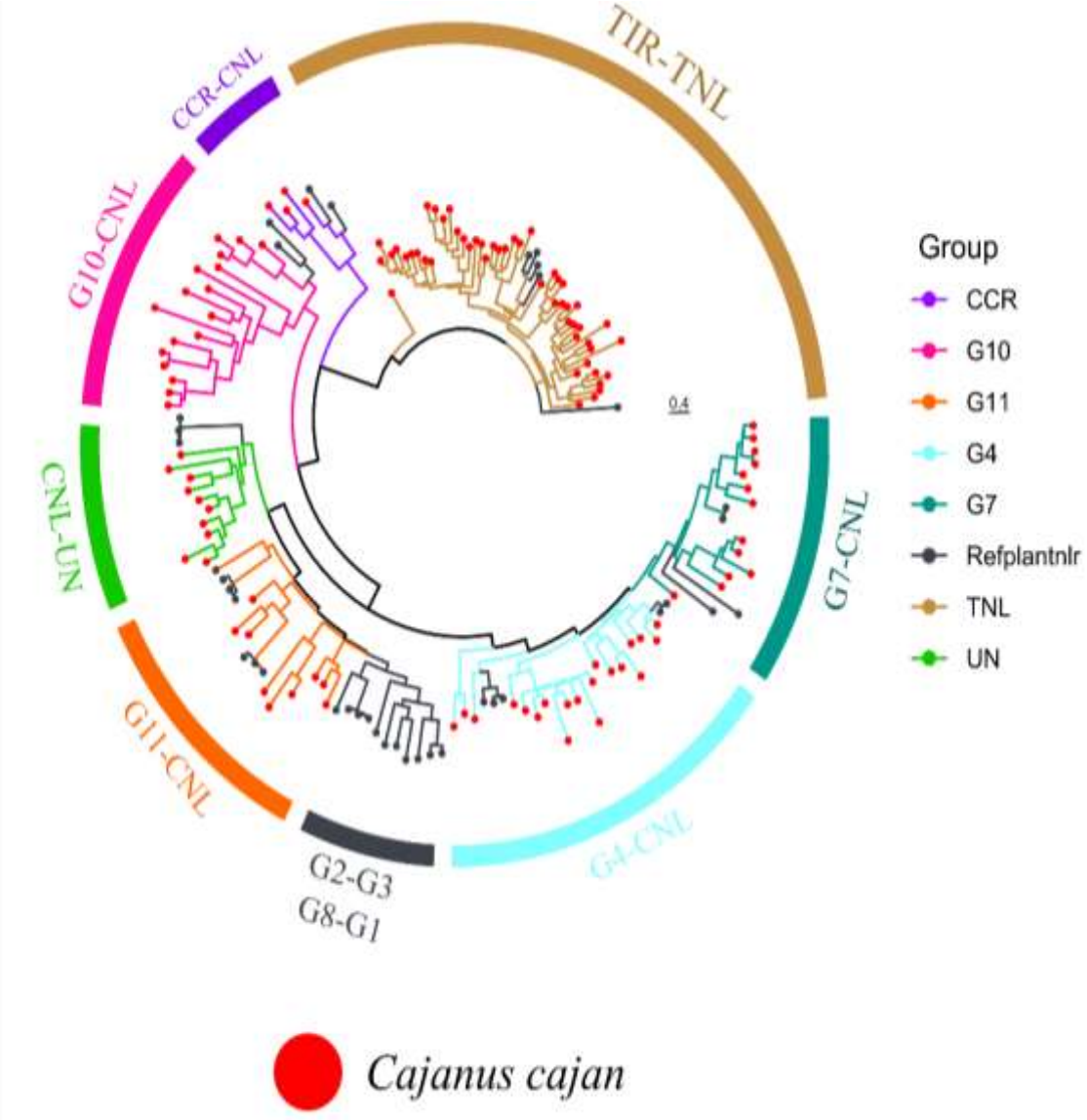

Fig. 6: Phylogenetic reconstruction of C. cajan NLR genes

Birth and Death of NLR Genes

Gene gain–loss analysis using CAFÉ revealed dynamic NLR gene evolution in C. cajan. Among the examined species, only 11 gene clusters were conserved indicating reduced overall conservation across the genus. Clade-specific expansion was evident suggesting that NLR gene diversification occurred independently within certain lineages.

After C. cajan diverged from its common ancestor, both gene birth and death rates increased significantly implying adaptive evolution driven by pathogen pressure. While Arabidopsis thaliana exhibited overall contraction of its NLRome, C. cajan demonstrated net gene gain and frequent terminal duplications (fig. 7).

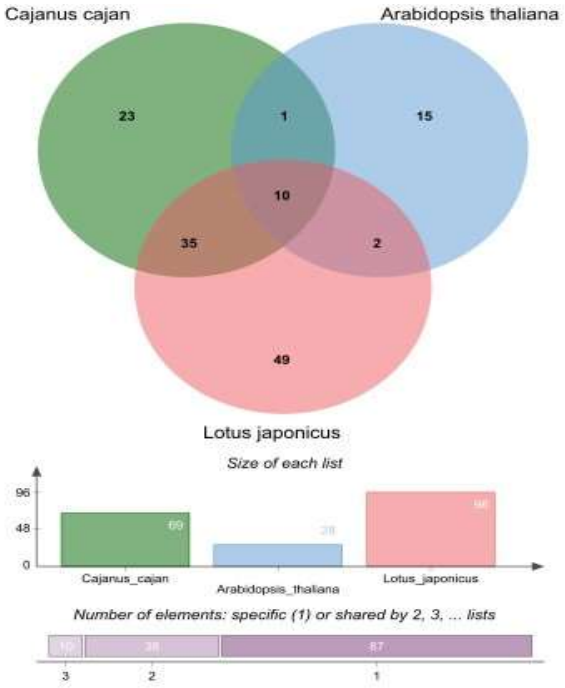

Fig. 7 Gene gain and loss analysis of C. cajan NLRs

Duplication History

Synonymous substitution rate (Ks) analysis revealed that C. cajan NLR duplication events occurred approximately 18 million years ago (mya), following its divergence from Glycine. The duplication pattern differed among subgroups: TIR-NLR, G4-NLR, G7-NLR and CCG10-NLR showed significant post-speciation duplication, while G11, CNL-Un and CCR-NLR exhibited reduced duplication activity (fig. 8). This indicates differential evolutionary pressures acting across the NLR subfamilies.

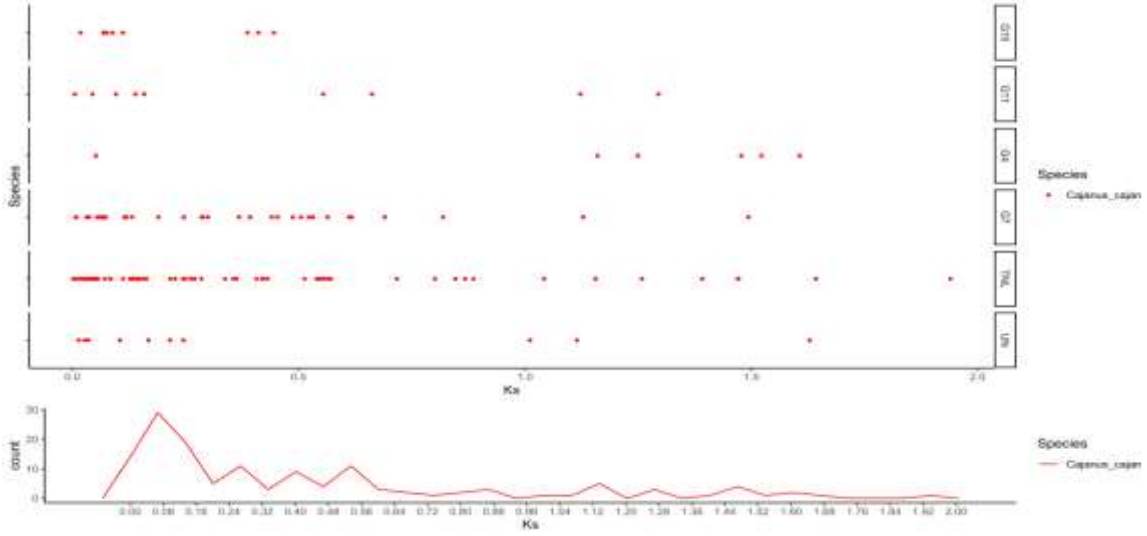

Fig. 8 Duplication history of C. cajan NLRs based on Ks values

Selection Pressure

Selection analysis using the Ka/Ks ratio demonstrated that most C. cajan NLR genes experienced purifying selection (Ka/Ks < 1), maintaining conserved functional domains. However, specific subfamilies exhibited signals of positive selection (Ka/Ks >1), implying adaptive diversification in response to pathogen coevolution.

Discussion

This study presents the first comprehensive genome-wide identification, characterization and evolutionary assessment of NLR genes in Cajanus cajan. A total of 205 NLR homologs were identified, consistent with other legume genomes confirming a well-developed immune gene repertoire. The predominance of CC and

TIR-type NLRs highlights their significant roles in the species' defense mechanisms, supporting previous findings by **Gururani et al. (2012)** who reported that angiosperms typically encode hundreds of NBS-LRR genes for broad-spectrum resistance.

Comparative domain and phylogenetic analyses revealed dynamic evolution within C. cajan NLRs. The monophyletic grouping of CC and CCG10-NLRs suggests ancient diversification events, while the polyphyletic structure of TNLs indicates ongoing expansion. The absence of G1–G8 CNL groups aligns with patterns observed in Cicer and other Fabaceae members implying a lineage-specific NLR evolution distinct from Solanaceae species.

The observed gene gain and loss dynamics suggest a strong influence of host–pathogen interactions on NLRome diversification. Consistent with the birth-and-death model of NLR evolution, C. cajan shows both clade-specific expansion and differential gene retention, a process previously described by **Shao et al. (2016)**. Moreover, the Ks-based duplication analysis (~18 mya) supports the idea that major NLR expansions occurred after specie identification, leading to adaptive diversification within Fabaceae.

Selection pressure analysis reinforces the notion that while most NLRs are conserved under purifying selection, subsets are under positive selection, consistent with report by **Zhang et al. (2022)** reflecting their ongoing adaptation to rapidly evolving pathogen populations. The presence of highly conserved FLY motifs and structural stability within NBARC domains further emphasizes the functional significance of these genes in maintaining robust immune signaling.

Overall, this study provides valuable insights into the genomic organization, evolution and adaptive diversification of the NLR gene family in C. cajan. These findings not only enhance understanding of legume immune evolution but also offer a genomic foundation for future resistance breeding and NLR engineering efforts in pigeon pea and related crops.

Conclusion

The exploration of NLR genes in Cajanus cajan provides valuable insights into the genetic foundation of disease resistance in this vital legume crop. The identification and classification of 205 NLR homologs reveal a complex yet well-organized immune gene family that has evolved through duplication and diversification to counter diverse pathogens. These results underscore the evolutionary adaptability of C. cajan in facing biotic challenges particularly from pests such as Helicoverpa armigera. Beyond their biological relevance, these findings offer a genomic framework for targeted breeding strategies and functional studies aimed at improving stress resilience and productivity in pigeon pea. Future investigations focusing on the expression profiling and functional validation of these candidate genes will further enhance our understanding of host–pathogen interactions in legumes.

References

Edgar, R. C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics, 5(1), 113.

Gururani, M. A., Venkatesh, J., Upadhyaya, C. P., Nookaraju, A., Pandey, S. K., & Park, S. W. (2012). Plant disease resistance genes: current status and future directions. Physiological and molecular plant pathology, 78, 51-65.

Hung, J. H., & Weng, Z. (2016). Sequence alignment and homology search with BLAST and ClustalW. Cold Spring Harbor Protocols, 2016(11), pdb-prot093088.

Keerthi, M. C., Manjunatha, L., Bhargavi, H. A., Mahesha, H. S., Puyam, A., & Dasgupta, D. (2022). Development of biotic-stress resistant pigeonpea.

In Genomic designing for biotic stress resistant pulse crops (pp. 309-351). Cham: Springer International Publishing.

Kourelis, J., Sakai, T., Adachi, H., & Kamoun, S. (2021). RefPlantNLR is a comprehensive collection of experimentally validated plant disease resistance proteins from the NLR family. PLoS Biology, 19(10), e3001124.

Kushwaha, A., & Mehta, C. M. (2023). Pigeon Pea (Cajanus cajan L.) Based Intercropping System: A Review. International Journal of Plant & Soil Science, 35(18), 1674-1689.

Madden, T. (2013). The BLAST sequence analysis tool. The NCBI handbook, 2(5), 425-436.

Marone, D., Russo, M. A., Laidò, G., De Leonardis, A. M., & Mastrangelo, A. M. (2013). Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. International journal of molecular sciences, 14(4), 7302-7326.

Prasad, D. V., Madhusudanan, S., & Jaganathan, S. (2015). uCLUST-a new algorithm for clustering unstructured data. ARPN Journal of Engineering and Applied Sciences, 10(5), 2108-2117.

Santos, M. D. L., Resende, M. L. V. D., Alves, G. S. C., Huguet-Tapia, J. C., Resende, M. F. R. D. J., & Brawner, J. T. (2022). Genome-wide identification, characterization, and comparative analysis of NLR resistance genes in Coffea spp. Frontiers in Plant Science, 13, 868581.

Shao ZhuQing, S.Z., Xue JiaYu, X.J., Wu Ping, W.P., Zhang YanMei, Z.Y., Wu Yue, W.Y., Hang YueYu, H.Y., Wang Bin, W.B. and Chen JianQun, C.J. (2016). Large-scale analyses of angiosperm nucleotide-binding site-leucine-rich repeat genes reveal three anciently diverged classes with distinct evolutionary patterns.

Singer, S. D., Laurie, J. D., Bilichak, A., Kumar, S., & Singh, J. (2021). Genetic variation and unintended risk in the context of old and new breeding techniques. Critical Reviews in Plant Sciences, 40(1), 68-108.

Steele, J. F., Hughes, R. K., & Banfield, M. J. (2019). Structural and biochemical studies of an NB-ARC domain from a plant NLR immune receptor. PloS one, 14(8), e0221226.

Xiao, S. Z., Armit, C., Edmunds, S., Goodman, L., Li, P., Tuli, M. A., & Hunter, C. I. (2019). Increased interactivity and improvements to the GigaScience database, GigaDB. Database, 2019, baz016.

Zhang, X., Liu, B., Zou, F., Shen, D., Yin, Z., Wang, R., He, F., Wang, Y., Tyler, B.M., Fan, W. and Qian, W. (2019). Whole genome re-sequencing reveals natural variation and adaptive evolution of Phytophthora sojae. Frontiers in Microbiology, 10, 2792.