Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The NBS-LRR Gene Family in Psidium guajava: Potential Applications of Genetic Markers for Germplasm Characterization and Disease Resistance

Layra de Medeiros Cardozo
Universidade Federal do Espírito Santo
Marcia Flores da Silva Ferreira
marcia.ferreira@ufes.br

Universidade Federal do Espírito Santo
Gilza Barcelos de Souza
Universidade Federal do Espírito Santo
Iana Pedro da Silva Quadros
Universidade Federal do Espírito Santo
Guadalupe del Rosario Quispe Saji
Universidade Federal do Espírito Santo
Giovanna Pinto Pires
Universidade Federal do Espírito Santo
Pedro Henrique Dias dos Santos
Louisiana State University of Alexandria
Adésio Ferreira
Universidade Federal do Espírito Santo

Research Article

Keywords: Biotic agents. Bioinformatics. Guava. Plant resistance. SSR

Posted Date: September 23rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10283625/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

The NBS-LRR Gene Family in *Psidium guajava*: Potential Applications of Genetic Markers for Germplasm Characterization and Disease Resistance

Layra de Medeiros Cardozo[1], Gilza Barcelos de Souza[1], Iana Pedro da Silva Quadros[1], Guadalupe del Rosario Quispe Saji[1], Giovanna Pinto Pires[1], Pedro Henrique Dias dos Santos[2], Adésio Ferreira[1], Marcia Flores da Silva Ferreira[1].

[1]Universidade Federal do Espírito Santo, Centro de Ciências Agrárias e Engenharias. Alto Universitário, S/N, Guararema, Alegre, ES, 29550-000, Brasil.

[2]Louisiana State University of Alexandria, College of Sciences, Technology, and Mathematics, 8100 Hwy 71 South Alexandria, LA 71302 USA.

Corresponding author: layrademedeiroscardozo@gmail.com; marcia.ferreira@ufes.br

Abstract

The NBS-LRR gene family plays a central role in plant immune responses; however, its characterization in *Psidium guajava*, a crop of global socioeconomic relevance, remains unexplored. In this study, 279 NBS-LRR genes were identified and classified into subclasses (CNL, TNL, RNL, and others) based on conserved domains. Comprehensive analyses were conducted to investigate gene structure, protein properties, subcellular localization, chromosomal distribution, conserved motifs, and domain organization. Phylogenetic analyses were conducted using full-length NBS-LRR protein sequences from *P. guajava*, *Arabidopsis thaliana*, and functionally validated nematode resistance genes revealed that several *P. guajava* proteins clustered with known nematode resistance genes. Notably, orthologs of *Mi-1*, *Gpa2*, and *Gro1-4* were identified, suggesting their potential role in nematode resistance. Furthermore, microsatellite analysis detected 25 SSRs within NBS-LRR protein-coding sequences, including two NBS-LRR genes, orthologous to *Mi-1*, *CaMi*, *Hero*, and *Gpa2* and *Cre1* and *Cre3*, respectively, containing microsatellite motifs. Variant calling analysis revealed a broad distribution of SNPs across the Psidium guajava genome, with NBS-LRR genes showing a high frequency of missense and synonymous variants, suggesting regions potentially associated with adaptation and pathogen resistance. These findings provide novel insights into the genetic basis of disease resistance in *P. guajava* and its interactions with pathogens. This study represents a significant step toward understanding the molecular mechanisms of pathogen

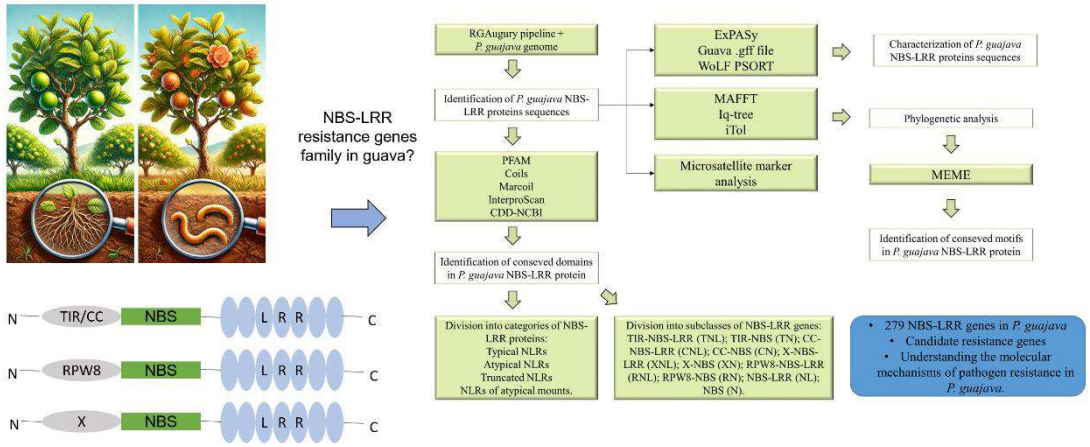
resistance in *P. guajava*, with implications for sustainable crop protection and productivity.

Keywords: Biotic agents. Bioinformatics. Guava. Plant resistance. SSR.

Highlights

- A total of 279 NBS-LRR genes were identified in *Psidium guajava*.
- Orthology between NBS-LRR proteins of guava and resistance proteins.
- Twenty-five SSR motifs were found within NBS-LRR gene sequences.
- Insights into *P. guajava*-pathogen interactions and candidate resistance genes.

Graphical abstract



1. Introduction

Plants face constant challenges from pests and pathogens (Coll et al. 2011). Through coevolution, they have developed complex and highly effective defense mechanisms, while pathogens have evolved strategies to overcome these defenses (Ramírez-Zavaleta et al. 2022). To counter biotic stress from pathogens, plants activate multiple defense mechanisms, primarily through two major systems: pathogen-triggered immunity (PTI) and effector-triggered immunity (ETI) (Gouveia et al. 2017).

Pathogen-triggered immunity (PTI) is initiated when pathogen- or microbe-associated molecular patterns (PAMPs or MAMPs) are recognized by pattern recognition

receptors (PRRs) on the surface of plant cells (Macho and Zipfel 2014; Gouveia et al. 2017). This recognition triggers a signaling cascade that reprograms gene expression and activates the production of specific defense molecules. PTI provides a rapid, broad-spectrum immune response against a wide range of pathogens (Hogenhout et al. 2009; Roux et al. 2014; Bigeard et al. 2015; Gouveia et al. 2017).

Effector-triggered immunity (ETI), in contrast, is mediated by the intracellular nucleotide-binding and leucine-rich repeat (NBS-LRR) receptors recognize specific pathogen effectors, either directly or indirectly (Jones and Dangl 2006; Tsuda and Katagiri 2010; Gouveia et al. 2017). This response often results in hypersensitivity reactions (HR), the production of reactive oxygen species (ROS), and salicylic acid (SA), as well as the expression of genes related to defense (Win et al. 2012; Gouveia et al. 2017). Although PTI and ETI share some downstream responses, they differ in timing and intensity. Together, these pathways orchestrate a robust transcriptional reprogramming that enhances plant defense (Ramírez-Zavaleta et al. 2022). In summary, the recognition of pathogens by PRRs and NBS-LRR contributes to the upregulation of genes involved in plant defense (Ramírez-Zavaleta et al. 2022). Notably, the NBS-LRR gene family plays a central role in ETI, making its members strong candidates for the identification of resistance genes.

NBS-LRR proteins are key components of ETI and are characterized by a modular structure composed of conserved domains. The N-terminal region typically contains either a Toll/interleukin-1 receptor (TIR) or a coiled-coil (CC) domain, both of which are involved in downstream signaling. The central region includes a nucleotide-binding site (NBS), also referred to as the NB-ARC domain, which possesses ATPase activity. The C-terminal region consists of leucine-rich repeats (LRRs), responsible for the recognition of pathogen-derived signals and, in some cases, may also participate in signaling (Inturrisi et al. 2020). Domain NBS and LRR have, respectively, ATPase activity and pathogen recognition (Inturrisi et al. 2020). These individual domains act together to provide effective immune responses against pathogens and pests (Bentham et al. 2020).

NBS-LRR proteins are generally classified into two major subgroups based on their N-terminal domains: TIR-NBS-LRR (TNL) proteins, which contain Toll/interleukin-1 receptor (TIR) domain, and non-TIR-NBS-LRR proteins, which contain a coiled-coil (CC), resistance to powdery mildew8 (RPW8), or another domain type (Meyers et al. 2003a; Xiao et al. 2010; Collier et al. 2011; Zhong et al. 2022). In addition to these

canonical forms, several atypical NBS-LRR proteins - characterized by truncated, fused, or rearranged domains - also play roles in plant immunity. These variant forms, have led to new classifications within this family: truncated NBS-LRR, atypical NBS-LRR and NBS-LRR with atypical assembly (Liu et al. 2015; Kapos et al. 2019).

The NBS-LRR gene family has been identified and characterized in various plant species, including both model organisms and economically important crops. For instance, it comprises 200 genes in *Arabidopsis thaliana* (Brassicaceae) (Meyers et al. 2003a), 535 in *Oryza sativa* (Poaceae) (Zhou et al. 2004), 438 in *Solanum tuberosum* (Solanaceae) (Jupe et al. 2012; Lozano et al. 2012), 355 in *Gossypium raimondii* (Malvaceae) (Wei et al. 2013), 228 in *Manihot esculenta* (Euphorbiaceae) (Lozano et al. 2015), and up to 2000 in *Triticum* species (Poaceae) (Gu et al. 2015). Other examples include 1,487 genes in *Raphanus sativus* (Brassicaceae) (Wang et al. 2017), 98 in *Musa acuminata* (Musaceae) (Chang et al. 2020), 167 in *Dioscorea rotundata* (Dioscoreaceae) (Zhang et al. 2020b), 289 in *Brassica juncea* (Brassicaceae) (Inturrisi et al. 2020), and 1,487 in *Eucalyptus* spp. (Myrtaceae) (Christie et al. 2016). Despite their relevance, the NBS-LRR gene family has not yet been studied in guava (*Psidium guajava*), highlighting a critical knowledge gap and an opportunity to explore potential resistance gene candidates.

P. guajava (Myrtaceae) holds significant economic value due to its high vitamin C content, which makes it valuable for the food industry, particularly in sweets, jams, and juices (Feng et al. 2021). Beyond its nutritional value, various parts of the plant exhibit notable medicinal properties, including anti-inflammatory, antiviral, and antimutagenic effects (Naseer et al. 2018). Brazil ranks as the sixth-largest guava producer worldwide, with an estimated annual production of 2.1 million tons in recent years (TRIDGE 2025). In 2023, guava production in Brazil generated approximately 1.31 billion reais in production value (IBGE 2020). However, guava production faces significant challenges from both abiotic and biotic stresses. Among the major biotic threats are anthracnose caused by *Colletotrichum gloeosporioides* (Nascimento et al. 2020), rust disease caused by *Puccinia psidii* (Ribeiro and Pommer 2002), bacterial blight caused by *Erwinia psidii* (Hermenegildo et al. 2019), and infestations by root-knot nematodes (*Meloidogyne* spp.) infestations (Rahaman Khan et al. 2022).

Root-knot nematodes, especially *Meloidogyne* species, represent a major threat, causing the formation of galls, which lead to substantial economic losses. These

nematodes are responsible for global crop losses estimated in billions of dollars annually (Sikandar et al. 2020). Currently, control strategies rely predominantly on chemical treatments, raising significant environmental concerns (Chen et al. 2007). Resistance genes against root-knot nematodes have been identified in several crops, most of which belong to the NBS-LRR gene family. Notable examples include *Mi-1.2* in tomato (Milligan et al. 1998), *Hero A* in potato (Bakker et al. 2006), *Ma* in plum (Claverie et al. 2011), and *Me* genes in pepper (Djian-Caporalino et al. 2001). These genes provide valuable references for the identification of candidate resistance genes in guava.

The emergence of DNA-based molecular markers has transformed the study of genetic diversity in cultivated plants, enabling faster and more accurate analyses. Among these markers, microsatellites — or SSRs (Simple Sequence Repeats) — stand out due to their codominant nature, multiallelic characteristics, and wide distribution across the genome, making them highly informative for genetic investigations (Savadi et al. 2021). SSRs exhibit high levels of polymorphism and have been extensively employed in studies of genetic diversity, genetic map construction, and marker-assisted breeding programs (Varshney et al. 2007). When located in genomic regions associated with functionally important genes, such as those from the NBS-LRR family, SSRs can reflect genetic variations related to pathogen resistance, and serve as strategic tools for assisted selection, as has been demonstrated in several crops (Quinton-Tulloch and Agronomy, Steele 2021; Bashir et al. 2022; Zhao et al. 2024; Alizadeh-Moghaddam et al. 2024).

Understanding the genetic and molecular mechanisms underlying guava's defense response to pathogens is essential. This study focuses on the identification and characterization of candidate NBS-LRR genes potentially involved in resistance to nematodes and other pathogens in *P. guajava*. Such advancements would support sustainable agricultural practices and help address food security challenges by producing more resilient crop varieties (Bentham et al. 2020). In this study, we present the first identification, characterization, and classification of NBS-LRR genes in *P. guajava*, along with phylogenetic, conserved motif, and microsatellite marker analyses, providing a foundation for future resistance gene discovery and breeding efforts.

2. Materials and methods

2.1. Annotation, *in silico* characterization, and classification of NBS-LRR genes in *P. guajava*

NBS-LRR genes in *P. guajava* were identified using the RGAugury pipeline (Li et al. 2016) applied to the *P. guajava* genome assembly GCA_016432845.1 (Feng et al. 2021). The presence of the characteristic NBS domain (PF00931) and additional domains was confirmed using PFAM v31.0 (Finn et al. 2014). These domains included TIR (PF01582), RX-N (PF18052), RPW8 (PF05659), and LRR (PF00560, PF07723, PF07725, and PF12799). The coiled-coil CC domain was validated through the Coils (Lupas et al. 1991) and Marcoil (Delorenzi and Speed 2002) programs, applying a 90% probability threshold. Further confirmation was performed via InterProScan (Mitchell et al. 2019) and the Conserved Domain Database CDD at NCBI (Lu et al. 2020).

The identified NBS-LRR genes were categorized to their domain architectures into four groups: typical NBS-LRR (with CC or TIR, NBS, and LRR domains), truncated NBS-LRR (missing CC or TIR domains), atypical NBS-LRR (featuring domains like Ank2, PGG, HMA, or Rx_N), and those with non-standard domain assembly, where domain order differs from typical NBS-LRR gene family members (Liu et al. 2015).

To characterize the NBS-LRR proteins, we used ExPASy's compute pI/Mw tool to calculate amino acid sequence length, theoretical isoelectric point (pI), and molecular weight (kDa) (Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., ... & Stockinger 2012). Chromosomal localization and nucleotide sequences, along with their base-pair lengths, were extracted from the *P. guajava* genome annotation file (.gff) (Feng et al. 2021). The distribution of NBS-LRR genes across chromosomes was visualized using Gene2Chrom v2.0 (Jiangtao et al. 2015). Subcellular localization predictions for the NBS-LRR proteins were performed with WoLF PSORT (Horton et al. 2007).

2.2. Phylogenetic analysis

Complete NBS-LRR protein sequences from *Arabidopsis thaliana* (Meyers et al. 2003a) as well as functionally characterized nematode-resistance genes identified in

various plant species (Vos et al. 1998; Van Der Vossen et al. 2000; Ernst et al. 2002; De Majnik et al. 2003; Paal et al. 2004; Chen et al. 2007; Xiao et al. 2010; Claverie et al. 2011; Saucet et al. 2016) were aligned with *P. guajava* NBS-LRR proteins using MAFFT (Katoh et al. 2002) with Blosum62 and 20 threads. The resulting alignments were used to construct a maximum-likelihood (ML) phylogenetic tree in IQ-TREE with 1000 bootstraps replicates (Nguyen and Evolution, Schmidt 2015), and the tree was visualized and edited using iTOL (Letunic and Bork 2021).

Reference sequences of nematode-resistant NBS-LRR proteins were obtained from the literature and databases, including NCBI (https://www.ncbi.nlm.nih.gov/) and the Plant Resistance Genes database (PRGdb) (http://prgdb.org).

2.3. Structural analysis of conserved motifs

Based on the phylogenetic analysis, proteins that clustered with known nematode-resistance proteins were selected for structural analysis of conserved motifs. The MEME Suite (https://meme-suite.org/meme/tools/meme) (Timothy L. Bailey et al. 2009) was used to predict conserved motifs, using the following parameters: any number of repetitions, a maximum of 20 motifs, and optimal motif length between 6 and 200 amino acid residues. To further characterize these motifs, they were compared with those previously described in *Arabidopsis thaliana* (Meyers et al. 2003a) and *Olea europaea* (Bettaieb and Bouktila 2020).

2.4. Identification of microsatellite markers in NBS-LRR Genes of *P. guajava*.

Genotypic data previously generated by Grossi et al. (2021) were used, obtained through DArTseq™ technology (Diversity Arrays Technology, Canberra, Australia). In the original study, DNA from 61 samples belonging to *Psidium guajava* genotypes (Supplementary Material 7) was extracted from leaves using the CTAB method (Doyle 1990) and subsequently sent to Diversity Arrays Technology for genomic library construction, sequencing on the Illumina HiSeq2500 platform, and identification of SNP markers (Kilian et al. 2012; Sansaloni et al. 2020). The sequences were processed using the proprietary DArTsoft-seq14 pipeline, and only markers with a call rate ≥ 0.7 and reproducibility ≥ 0.95 were retained (Grossi et al. 2021).

Based on these genotypic data, the microsatellite analyses described in this study were performed.

MISA, a well-established PERL5-script tool for the identification of MIcroSAtellite markers in DNA sequences (Beier et al. 2017), was used with default parameters to analyze the coding sequences of *P. guajava*. The output was processed using an R script that incorporated the following packages: xlsx (https://github.com/colearendt/xlsx), dplyr (https://github.com/tidyverse/dplyr), showtext (https://github.com/yixuan/showtext), forcats (https://github.com/tidyverse/forcats), and ggplot2 (https://github.com/tidyverse/) to generate Figure 9.

2.5. Variant calling in NBS-LRR genes

2.5.1. Processing of raw reads

Initially, the reads were trimmed using Trimmomatic SE (Bolger et al. 2014) to remove adapters with a maximum mismatch of 2, a quality score cutoff of 30, and 10 removal attempts. Additionally, initial and final bases with a quality score below 20 were removed, sequences were trimmed when the average quality within a 4-base window dropped below 20 and reads shorter than 50 nucleotides were discarded. Subsequently, they were deduplicated using fastp (Chen et al. 2018) excluding bases with a quality score below 20 and discarding reads shorter than 50 nucleotides. Read quality was assessed using FastQC (Andrews 2010) and MultiQC (Ewels et al. 2016).

2.5.2. Mapping reads to the reference genome

BWA-MEM2 (Li and Durbin 2009; Vasimuddin et al. 2019) was used to index the reference genome of *Psidium guajava* (Feng et al. 2021) from the NCBI (GCA_016432845.1). The processed reads were aligned and mapped to the reference genome using BWA-MEM2. The resulting SAM file was converted to BAM format using SAMtools (Li and Durbin 2009; Danecek et al. 2021). Using Picard (Broad Institute 2019), BAM files were sorted based on read position, filtered to remove redundant reads, and annotated to identify read groups for variant calling.

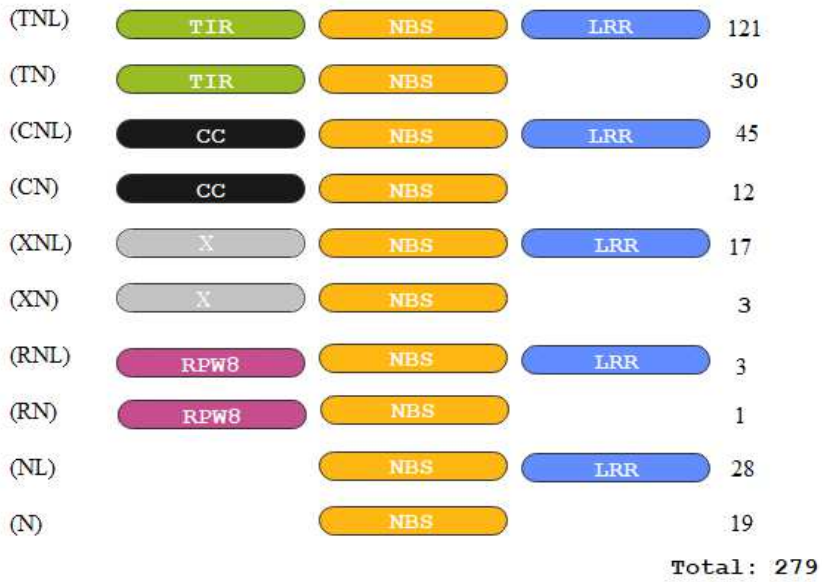
2.5.3.Variant Calling and Annotation

The processed BAM files were piled up to generate coverage and depth profiles for each genomic position and, based on these profiles, identify genomic variants using BCFtools (Li 2011; Danecek et al. 2021). Descriptive statistics of the generated VCF files were calculated using RTG-tools (Cleary et al. 2015). Variants were classified as either InDels or SNPs using VCFtools (Danecek et al. 2021). To ensure variant quality, SNPs were subjected to rigorous filtering criteria, including: the presence of the alternative allele in at least one sample, detection of the SNP in at least 90% of the samples, and a minimum Phred quality score of 20 — also using VCFtools. The quality of filtered variants was assessed using the VcfR package (Knaus and Grünwald 2017). Variant impact was annotated using SnpEff (Cingolani et al. 2012) with a GFF genomic annotation file predicted by Augustus (Stanke et al. 2004), using the Arabidopsis thaliana genome as the reference.The data was processed in R (R Core Team 2024b) v. 4.4.1, within R Studio (Posit Team 2024) v. 2024.12.0+467, using the packages patchwork (Pedersen 2025), showtext (Qiu 2024) e tidyverse (Wickham et al. 2019).

3. Results

3.1.Identification and Classification of NBS-LRR Genes in *Psidium guajava.*

A total of 279 NBS-LRR genes were identified in the *P. guajava* genome. Most of the predicted proteins (246) were classified as “Disease Resistance Protein, Plants” (IPR044974), while the remaining proteins were distributed across seven additional protein families (Table S1). Key conserved domains detected included NBS (PF00931), TIR (PF01582), RX-N (PF18052), RPW8 (PF05659), and various LRR domains (PF00560, PF07723, PF07725, PF12799). Based on the presence and combination of these domains, the proteins were categorized into ten subclasses: TIR-NBS-LRR (TNL), TIR-NBS (TN), CC-NBS-LRR (CNL), CC-NBS (CN), X-NBS-LRR (XNL), X-NBS (XN), RPW8-NBS-LRR (RNL), RPW8-NBS (RN), NBS-LRR (NL), and NBS (N) (Figure 1; Supplementary Material 1 (Table 3).



264

265 Figure 1. Subclassification of the 279 proteins encoded by the NBS-LR family genes in
266 *Psidium guajava*.

267 Further, the proteins were grouped into three categories: typical NBS-LRRs (17
268 CNL, 90 TNL, and 3 RNL), truncated forms (100 proteins with partial domain
269 composition), and proteins with atypical domain assemblies, which included additional
270 domains such as FolB_dom, CHASE, AP2, and more (Table 1).

271 Table 1. Classification of NBS-LRR proteins in *Psidium guajava*

Subclass	Typical NBS-LRR	Atypical NBS-LRR	NBS-LRR of atypical mounts	TruncatedNBS-LRR
CNL (45)	CC-NB-LRR (17)	CC-NB-LRR-FolB_dom (1),CC-NB-NB-LRR-AP2 (1)	X-CC-NB-LRR (23), X-CC-NB-NB-LRR (2), X-X-CC-TIR-NB-NB-LRR (1)	-
CN (12)	-	X-CC-NB-CHASE (1)	-	CC-NB (6), X-CC-NB (3), X-CC-NB-NB (1), X-CC-TIR-NB (1)
TNL (121)	TIR-NB-LRR (90)	CC-TIR-NB-LRR-HSF_DNA-bd (1), CC-TIR-NB-LRR-Oleasin (1), TIR-NB-LRR-	TIR-CC-NB-LRR (9), CC-TIR-TIR-NB-NB-LRR (1), TIR-NB-NB-	-

		CMP_dCMP_dom (1), TIR-NBS-LRR-Prot_kinase_dom (2), TIR-NB-LRR-VQ_motif (1), TIR-TIR-NB-LRR-C_JID (1)	LRR (5), TIR-TIR-NB-LRR (3), TIR-TIR-NB-NB-LRR (3), TIR-TIR-TIR-NB-NB-LRR (1), TIR-TIR-TIR-NB-NB-NB-LRR (1), X-CC-TIR-NB-LRR (1)	
TN (30)	-	TIR-NB-Jacalin-like_lectin_dom (2), TIR-NB-Trbl (1)	-	TIR-NB (26), TIR-TIR-NB-NB (1)
XNL (17)	-	X-NB-LRR-DUF2431 (1) X-NB-LRR-Pyridoxal-dependent decarboxylase conserved domain (1)	X-NB-NB-LRR (1)	X-NB-LRR (14)
XN (3)	-	X-NB-NB-Pribosyltran-Hsp70 (1)	-	X-NB (2)
RNL (3)	RPW8-NB-LRR (3)	-	-	-
RN (1)	-	-	-	RPW8-NB (1)
NL (28)	-	BED Zinc finger-NB-LRR (1) NB-LRR- Heavy-metal-associated domain (1)	-	NB-LRR (25) NB-NB-LRR (1)

TIR (Toll interleukin-1 receptor domain): Toll interleukin-1 receptor domain; CC (coiled coil domain): coiled coil domain; LRR (leucine-rich repeat domain): leucine-rich repeat domain; NB (nucleotide-binding ATPase domain): nucleotide-binding or hydrolysis domain; X: uncharacterized domain (in this study, corresponding to the Rx_N domain); RPW8 (Resistance to Powdery Mildew 8): resistance to powdery mildew domain; HMA (Heavy-metal-associated domain): heavy-metal-associated domain; Ank2 (ankyrin-B for broadly expressed): ankyrin B domain; PGG: undefined function domain. The number in parentheses indicates the quantity of proteins with these domain compositions listed in the table. Source: Adapted from Liu et al., 2015.

The characterization of NBS-LRR genes and their encoded proteins, including the number of amino acids (aa), protein molecular weight (kDa), theoretical isoelectric point

(pI), coding sequence length in base pairs (bp), subcellular localization, chromosomal location, and the presence, sizes, and positions of domains, is provided in the Supplementary Material 1 (Table 1). The corresponding coding sequences (CDS) and their translated amino acid sequences are provided in Supplementary Material 4 and 5, respectively.

As shown in Figure 2 and Supplementary Material 1 (Table 2), the amino acid sequences ranged from 86 (evm.model.ctg40.162) to 2,988 (evm.model.ctg17.321), with the greatest variation observed in the CNL subclass and the least in the RN, RNL, and XN subclasses. Molecular weights ranged from 9.72 kDa (evm.model.ctg40.162) to 337.30 kDa (evm.model.ctg17.321), again showing the highest variation in the CNL subclass and the smallest in the RN, RNL, and XN subclasses. Theoretical isoelectric points ranged from 4.47 pI (evm.model.ctg14.1138) to 9.38 pI (evm.model.ctg20.539), with a wide range of variation across subclasses except for RN, RNL, and XN. Coding sequence lengths in base pairs varied from 364 bp (evm.model.ctg40.162) to 118,706 bp (evm.model.ctg13.1522), showing the greatest variation in the TNL, CNL, and TN subclasses and the least in RN.

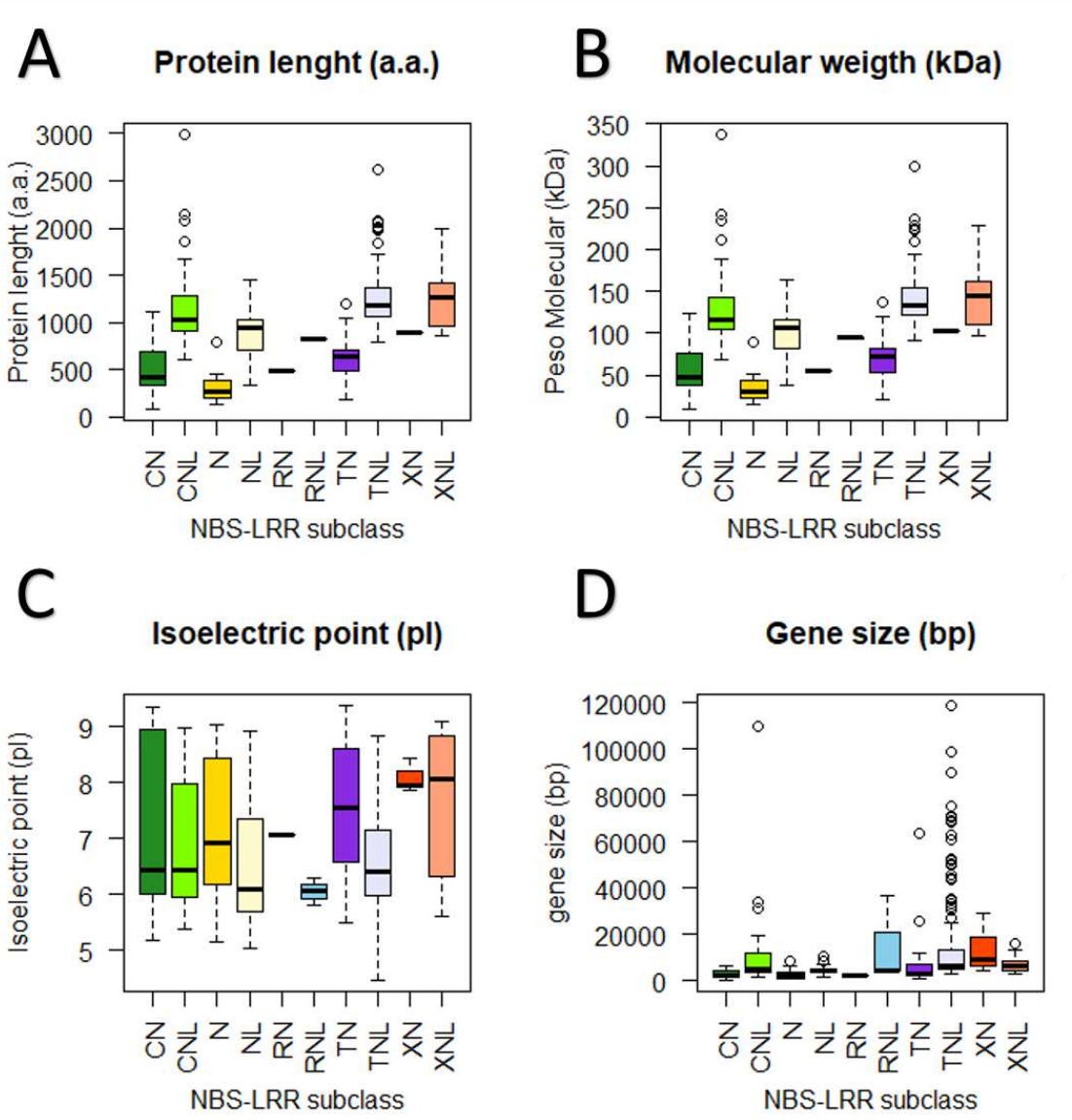


Figure 2. Chemical and structural characteristics of NBS-LRR proteins in guava. The y-axes represent the variables analyzed: Protein length (a.a.) (boxplot A), Molecular weight (kDa) (boxplot B), Isoelectric point (pI) (boxplot C), and Gene size (bp) (boxplot D), while the x-axis represents the subclasses of the NBS-LRR gene family (CN, CNL, N, NL, RN, RNL, TN, TNL, XN, and XNL). Colored bar charts display the average variation of each characteristic of NBS-LRR proteins within each subclass. All plots were generated using the R software.

Subcellular localization analysis revealed that most NBS-LRR proteins were nuclear, with subclass-specific distributions across other organelles (Table 2). Subcellular

localization predictions placed most proteins in the nucleus, reflecting its likely role in transcriptional regulation and pathogen recognition.

Table 2. Predicted subcellular localization of NBS-LRR proteins in guava. The number of proteins of each subclass is indicated, along with the total number of NBS-LRR proteins predicted in each organelle.

Organelle	**TNL**	**TN**	**CNL**	**CN**	**RNL**	**RN**	**NL**	**N**	**XNL**	**XN**	**Total NBS-LRR**
Cytoplasm	21	7	5	6	1	1	7	10	3	1	62
Chloroplast	27	11	10	3			5	3	4	1	64
Plasma membrane	10		2	2							15
Mitochondria	2		1						1		3
Nucleus	57	10	25	1	2		16	6	8	1	126
Peroxisome	3	2	1						1		7
Endoplasmic reticulus			1								1
Vacuole	1										1

Chromosomal mapping showed genes were unevenly distributed, with the highest density on chromosomes 3 (61 genes) and 8 (56 genes), and fewer genes on chromosomes 1 (5 genes) and 9 (7 genes) (Figure 3; Supplementary Material 1 – Table 1).

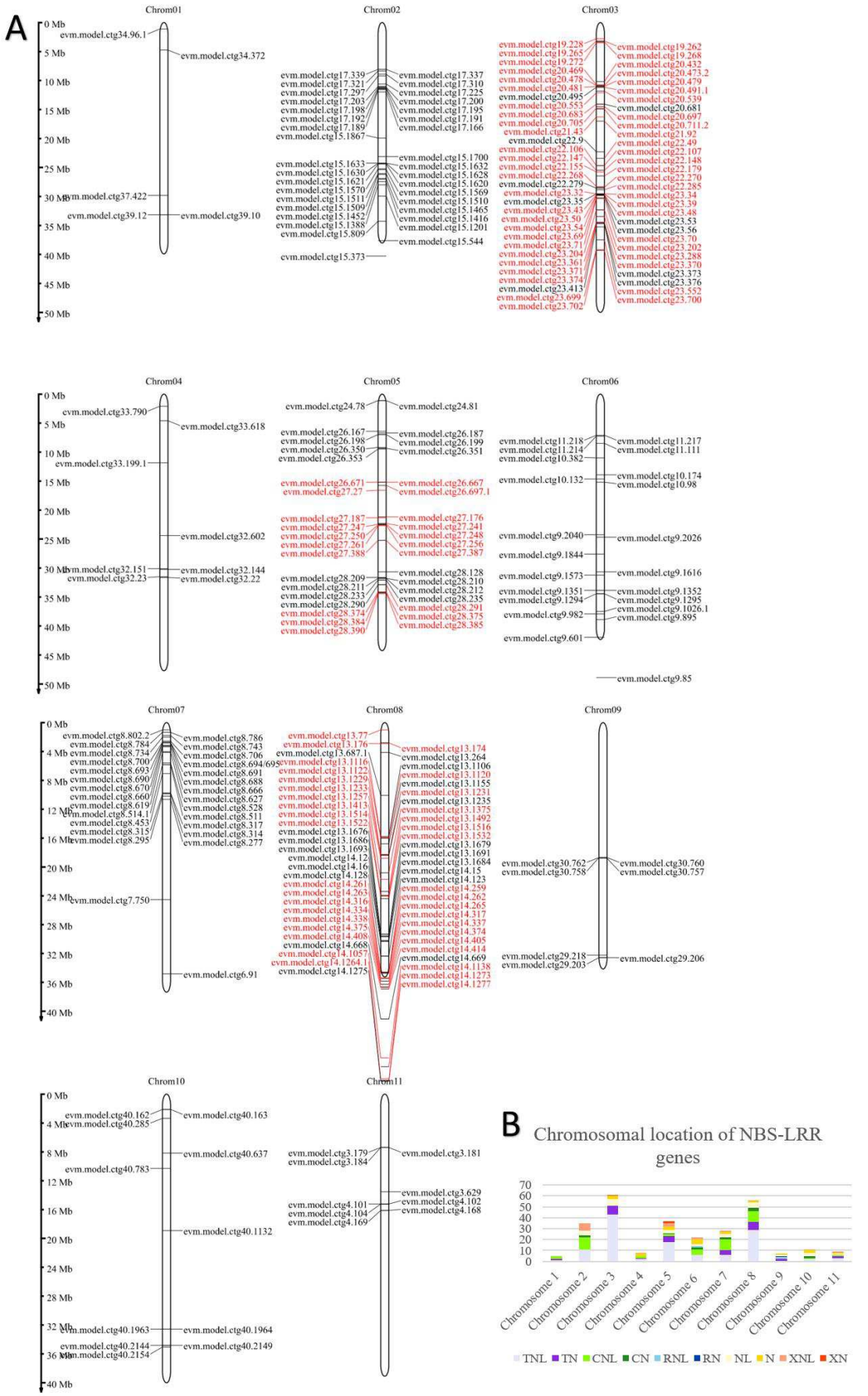

A
Chrom01
Chrom02
Chrom03
Chrom04
Chrom05
Chrom06
Chrom07
Chrom08
Chrom09
Chrom10
Chrom11
B
Chromosomal location of NBS-LRR genes
Chromosome 1
Chromosome 2
Chromosome 3
Chromosome 4
Chromosome 5
Chromosome 6
Chromosome 7
Chromosome 8
Chromosome 9
Chromosome 10
Chromosome 11
TNL TN CNL CN RNL RN NL N XNL XN

Figure 3. Chromosomal mapping (A) and graphical representation of the location (B) of NBS-LRR genes in *Psidium guajava*. Genes highlighted in red correspond to TIR-NBS-LRR subgroup members.

3.2.Phylogenetic relationships reveal evolutionary proximity and orthology between NBS-LRR proteins of guava and resistance proteins.

To investigate evolutionary relationships and potential orthology, a phylogenetic tree was constructed using the full-length NBS-LRR protein sequences from *P. guajava*, *A. thaliana*, and functionally characterized nematode resistance (R) proteins (Figure 4-A).

Three main clades were identified: the first separating the evm.model.ctg32.602 (TN) protein, the second separating the evm.model.ctg10.132 (CN) protein, and the third grouping the remaining proteins into multiple clades. The third branch is subdivided into two subgroups. The first subgroup includes proteins from the RN and RNL subclasses, along with most CN, CNL, XNL, and XN proteins. The second subgroup predominantly contains TN and TNL proteins. Below, the primary and significant clades formed are listed.

The AAP44390.1 protein (subclass: TNL, encoding gene: *Gro1-4*, host: *Solanum tuberosum*, pathogen: *Globodera rostochiensis*) (Paal et al. 2004) clustered with the guava protein evm.model.ctg40.285 (typical TNL) (Figure 4-B).

Proteins AAC97933.1 (subclass: CNL, encoding gene: *Mi-1*, host: *Solanum lycopersicum*, pathogens: *Meloidogyne javanica*, *Meloidogyne incognita*, and *Meloidogyne arenaria*) (Vos et al. 1998), ABE68835.1 (subclass: NL, encoding gene: *CaMi*, host: *Capsicum annuum*, pathogen: *M. incognita*) (Chen et al. 2007), CAD29729.1 (subclass: CNL, encoding gene: *Hero*, host: *S. lycopersicum*, pathogens: *Globodera. pallida* and *G. rostochiensis*) (Ernst et al. 2002), and AAF04603.1 (subclass: CNL, encoding gene: *Gpa2*, host: *S. lycopersicum*, pathogen: *G. pallida*) (Van Der Vossen et al. 2000) formed a clade with the guava proteins evm.model.ctg15.1201 (XNL), evm.model.ctg9.85 (CNL), evm.model.ctg40.1964 (XNL), evm.model.ctg40.1963

(CNL), evm.model.ctg40.2149 (NL), evm.model.ctg13.1155 (NL), evm.model.ctg40.2154 (NL), and evm.model.ctg40.2144 (N) (Figure 4-C).

Proteins CAD12795.1 (subclass: NL, encoding gene: *Cre3*, host: *Triticum*, pathogen: *Heterodera avenae*) (De Majnik et al. 2003) and AAM94164.1 (subclass: NL, encoding gene: *Cre1*, host: *Triticum*, pathogen: *H. avenae*) (De Majnik et al. 2003) clustered with guava proteins evm.model.ctg17.337 (CNL), evm.model.ctg17.339 (CN), evm.model.ctg17.297 (XNL), evm.model.ctg17.321 (CNL), evm.model.ctg15.1416 (CNL), evm.model.ctg15.1452 (XNL), evm.model.ctg15.1388 (XNL), evm.model.ctg15.544 (XNL), evm.model.ctg13.687.1 (CNL), evm.model.ctg40.637 (CN), evm.model.ctg6.91 (XNL), and with *Arabidopsis* proteins AT3G14460 (CNL) and AT3G14470 (CNL) (Figure 4-D).

The protein ADP37705 (subclass: TNL, encoding gene: *Rpm2*, host: *Prunus sogdiana*, pathogen: *M. incognita*) (Xiao et al. 2010) and the *Arabidopsis* protein AT3G07040 (CNL) clustered with guava proteins evm.model.ctg14.669 (CNL), evm.model.ctg14.668 (XNL), evm.model.ctg37.422 (CNL), evm.model.ctg26.350 (XNL), evm.model.ctg26.353 (XNL), evm.model.ctg26.351 (XNL), evm.model.ctg32.23 (XNL), evm.model.ctg32.22 (CNL), and evm.model.ctg32.144 (CNL) (Figure 4-C).

The main clades of interest in this study are those comprising both nematode resistance proteins and guava proteins, suggesting evolutionary proximity and orthology. These results highlight promising candidate proteins for further investigation of plant disease resistance mechanisms. Moreover, the placement of *Arabidopsis* proteins within these clades reinforces the classification of guava NBS-LRR proteins based on conserved motifs.

Conversely, the CAR94518.1 protein (subclass: TNL, encoding gene: *Ma*, host: *Prunus spp.*, pathogen: *Meloidogyne spp.*) (Claverie et al. 2011; Saucet et al. 2016) showed significant divergence from the protein sequences analyzed in the phylogeny (Figure 4-B).

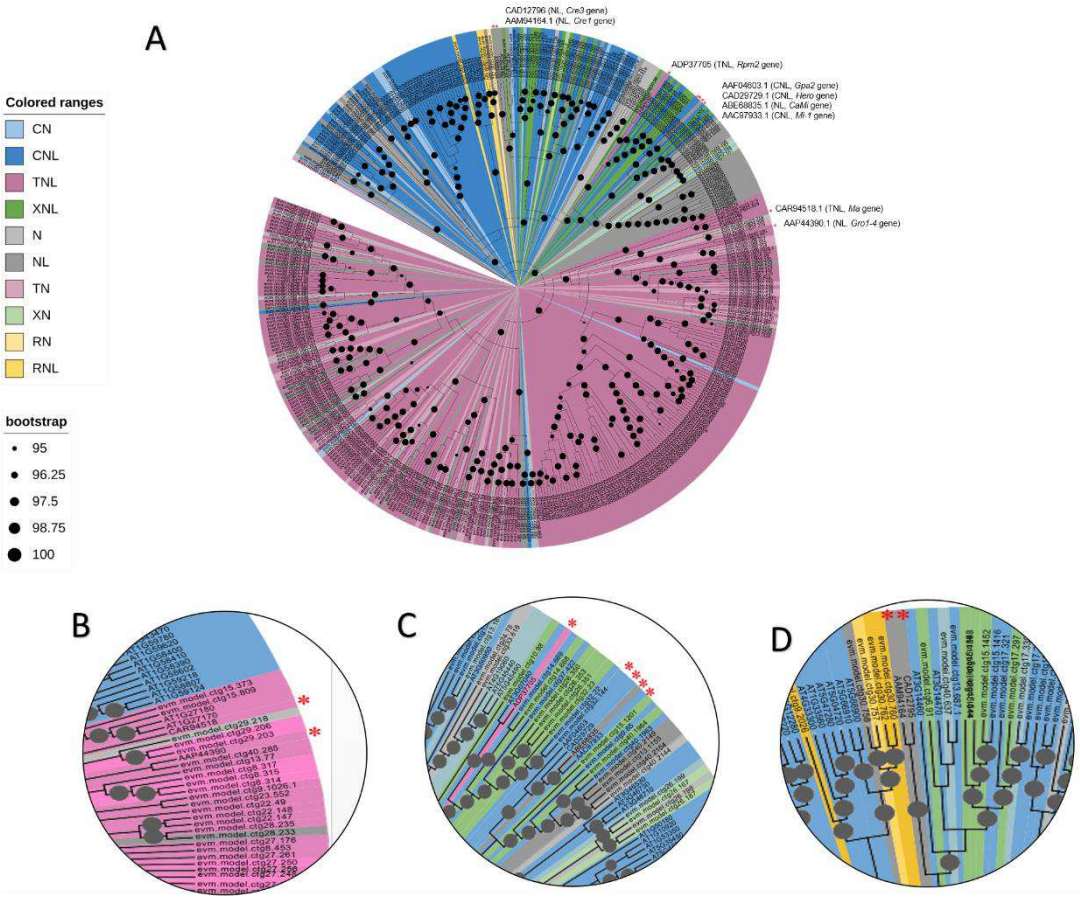


Figure 4. Phylogenetic analysis of NBS-LRR proteins from *Psidium guajava*, *Arabidopsis thaliana*, and nematode resistance genes. (A) Phylogenetic tree constructed from complete NBS-LRR protein sequences from *Psidium guajava* (279 sequences), *Arabidopsis thaliana* (154 sequences), and the proteins encoded by nematode resistance genes *Mi-1* (AAC97933.1), *Gpa2* (AAF04603.1), *Cre3* (CAD12795.1), *Cre1* (AAM94164.1), *Gro1-4* (AAP44390.1), *Hero* (CAD29729.1), *Ma* (CAR94518.1), *Rpm2* (ADP37705.1), and *CaMi* (ABE68835.1). Nematode resistance proteins are indicated by red asterisks. Bootstrap values and protein subclasses (color-coded) are detailed in the legend. The phylogeny was generated using IQ-TREE v1.6.12 with the maximum likelihood method and edited in iTOL. Panels B-D show highlighted clades within the phylogeny of *P. guajava* NBS-LRR proteins, *A. thaliana* proteins, and nematode resistance proteins.

3.3. Conserved Motifs Associated with Resistance Functions.

To investigate whether the NBS-LRR proteins forming phylogenetic clades with nematode resistance proteins share conserved motifs, amino acid sequences were analyzed using the MEME program, focusing on the NBS, TIR, and CC domains. A total of twenty putative conserved motifs were detected in each group of genes. To aid motif prediction, previously described motif sequences from *Arabidopsis thaliana* (Meyers et al. 2003a) and *Olea europaea* (Bettaieb and Bouktila 2020) were used for comparison (Supplementary Material 1 – Table 4).

The NBS-LRR proteins evm.model.ctg40.285 (TNL) from *P. guajava* and AAP44390.1 (TNL, *Gro1-4* gene), which confers resistance to *Globodera rostochiensis*, and formed the first highlighted clade, both exhibited conserved motifs including the P-loop and kinase-3 motifs (Figure 5).

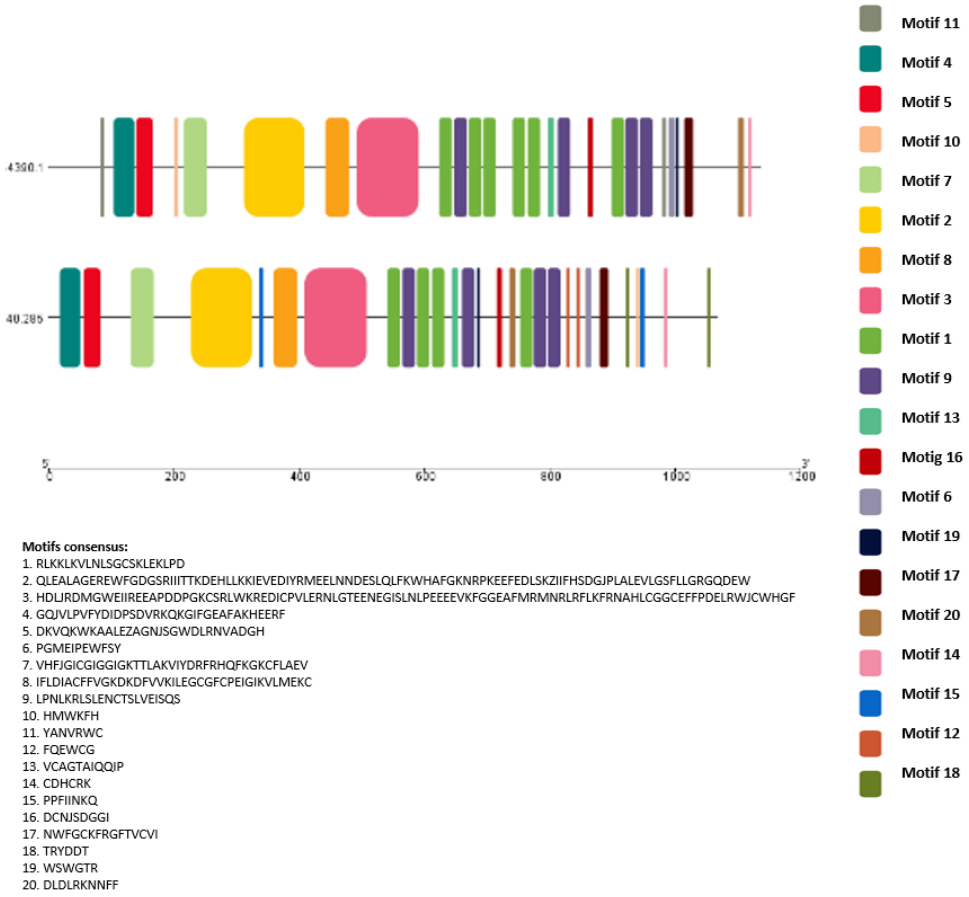
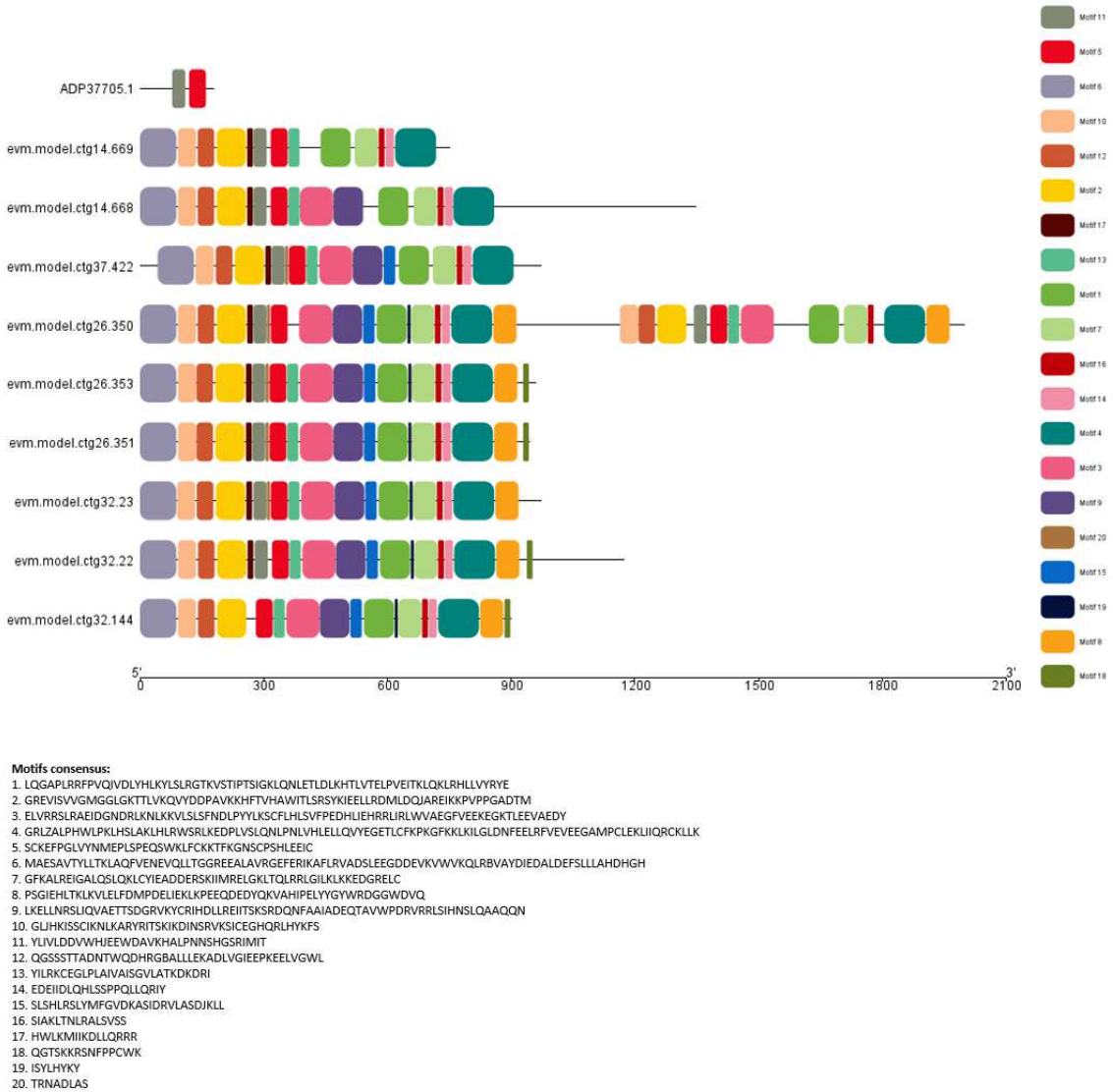


Figure 5. Twenty conserved motifs identified in the nematode resistance protein AAP44390.1 (TNL, *Gro1-4* gene) conferring resistance to *G. rostochiensis* and the NBS-LRR protein evm.model.ctg40.285 (TNL) from *P. guajava*.

The NBS-LRR proteins evm.model.ctg14.669 (CNL), evm.model.ctg14.668 (XNL), evm.model.ctg37.422 (CNL), evm.model.ctg26.350 (XNL), evm.model.ctg26.353 (XNL), evm.model.ctg26.351 (XNL), evm.model.ctg32.23 (XNL), evm.model.ctg32.22 (CNL), and evm.model.ctg32.144 (CNL) from *P. guajava*, as well as ADP37705 (TNL, *Rpm2* gene) conferring resistance to *M. incognita*, which formed the second highlighted clade, exhibited conserved motifs such as RNBS-A, Kinase-2, Kinase-3, L5, L6, P-loop, and C1 (Figure 6).



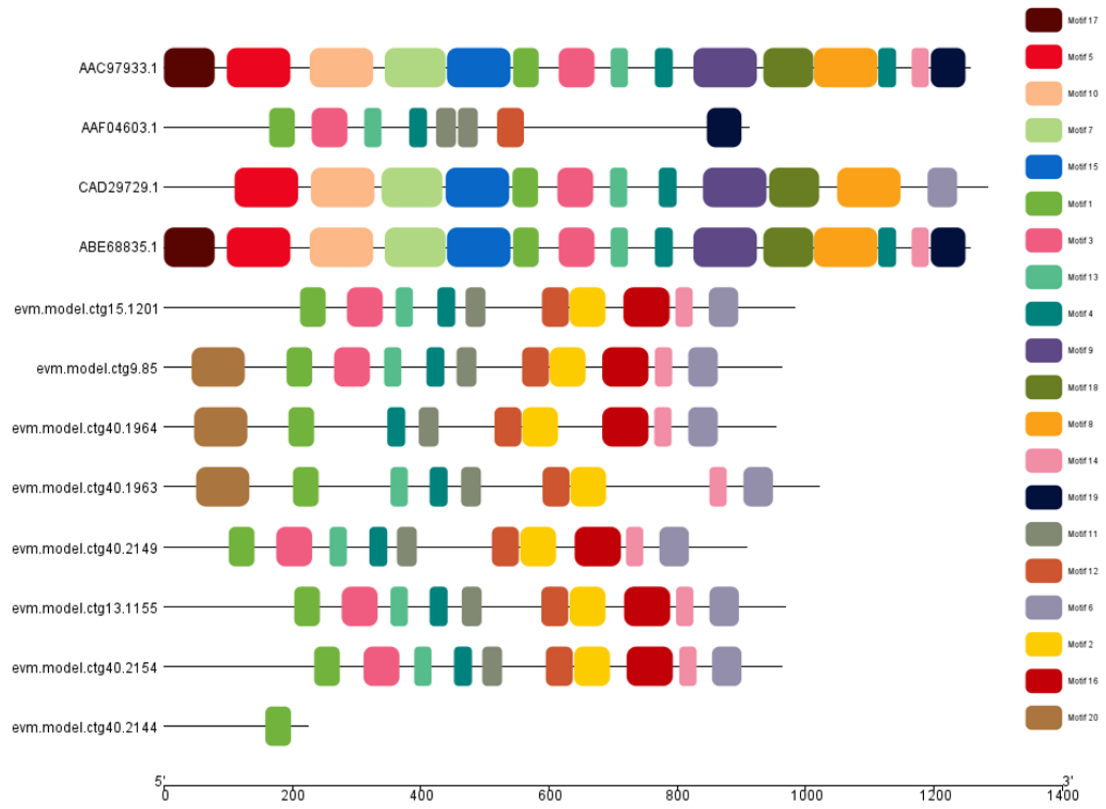
Motifs consensus:

1. LQGAPLRRFPVQIVDLYHLKYLSLRGTKVSTIPTSIGKLQNLETLDLKHTLVTELPVEITKLQKLRHLLVYRYE
2. GREVISVVGMGGLGKTTLVKQVYDDPAVKKHFTVHAWITLSRSYKIEELLRDMLDQJAREIKKPVPPGADTM
3. ELVRRSLRAEIDGNDRLKNLKKVLSLSFNDLPYYLKSCFLHLSVFPEDHLIEHRRLIRLWVAEGFVEEKEGKTLEEVAEDY
4. GRLZALPHWLPKLHSLAKLHLRWSRLKEDPLVSLQNLPNLVHLELLQVYEGETLCFKPKGFKKLKILGLDNFEELRFVEVEEGAMPCLEKLIIQRCKLLK
5. SCKEFPGLVYNMEPLSPEQSWKLFCKKTFKGNSCPSHLEEIC
6. MAESAVTYLLTKLAQFVENEVQLLTGGREEALAVRGEFERIKAFLRVADSLEEGDDEVKVWVKQLRBVAYDIEDALDEFSLLLAHDHGH
7. GFKALREIGALQSLQKLCYIEADDERSKIIMRELGKLTQLRRLGILKLKKEDGRELC
8. PSGIEHLTKLKVLELFDMPDELIEKLKPEEQDEDYQKVAHIPELYYGYWRDGGWDVQ
9. LKELLNRSLIQVAETTSDGRVKYCRIHDLLREIITSKSRDQNFAAIADEQTAVWPDRVRRLSIHNSLQAAQQN
10. GLJHKISSCIKNLKARYRITSKIKDINSRVKSICEGHQRLHYKFS
11. YLIVLDDVWHJEEWDAVKHALPNNSHGSRIMIT
12. QGSSSTTADNTWQDHRGBALLLEKADLVGIEEPKEELVGWL
13. YILRKCEGLPLAIVAISGVLATKDKDRI
14. EDEIIDLQHLSSPPQLLQRIY
15. SLSHLRSLYMFGVDKASIDRVLASDJKLL
16. SIAKLTNLRALSVSS
17. HWLKMIIKDLLQRRR
18. QGTSKKRSNFPPCWK
19. ISYLHYKY
20. TRNADLAS

Figure 6. Twenty conserved motifs of the ADP37705 protein (TNL, *Rpm2* gene) conferring resistance to *M. incognita* and the NBS-LRR proteins evm.model.ctg14.669 (CNL), evm.model.ctg14.668 (XNL), evm.model.ctg37.422 (CNL),

418 evm.model.ctg26.350 (XNL), evm.model.ctg26.353 (XNL), evm.model.ctg26.351
419 (XNL), evm.model.ctg32.23 (XNL), evm.model.ctg32.22 (CNL), and
420 evm.model.ctg32.144 (CNL) from *P. guajava*.

421 Meanwhile, the NBS-LRR proteins evm.model.ctg15.1201 (XNL),
422 evm.model.ctg9.85 (CNL), evm.model.ctg40.1964 (XNL), evm.model.ctg40.1963
423 (CNL), evm.model.ctg40.2149 (NL), evm.model.ctg13.1155 (NL),
424 evm.model.ctg40.2154 (NL), and evm.model.ctg40.2144 (N) from *P. guajava*;
425 AAC97933.1 (CNL, *Mi-1* gene) conferring resistance to *Meloidogyne spp*.; ABE68835.1
426 (CNL, *CaMi* gene) conferring resistance to *M. incognita*; CAD29729.1 (CNL, *Hero* gene)
427 conferring resistance to *G. pallida* and *G. rostochiensis*; and AAF04603.1 (CNL, *Gpa2*
428 gene) conferring resistance to *G. pallida*, which formed the third highlighted clade,
429 exhibited conserved motifs, including P-loop, RNBS-A, Kinase-2, Kinase-3, GLPL, and
430 MHDV (Figure 7).



Motifs consensus:

1. VISIVGMPGIGKTTLARSVYNDLEIKHHFDCRAWVTVSQEY
2. RYLGLRSTVLDSIPSSVGBLYYLETLDLKHTNITSLPSSIWKARNLRHLYLNEVYFD
3. RKALMGKRYLIVLDDVWDTETWDELTIPFPDAKKGSRIILTTREKEVALYVDPNTDP
4. PSHLKPCLLYFALFPKDSPIPTRRLVRLW
5. NMDDCISLYHRSYKSDAIMMDEQLDFLLLNLYHLSKHHAEKIFPGVTQYEVLQNVCGNIRDFHGLILNGCIKHEMVENVLPLFQLMAERVGHFLWEDQTD
6. FPSLRVLKLWKLEKLKKWTIEEGAMPILEELEIRSCGELTTIEGJKH
7. GNNQTNCATLDLLENIELLKKDLKHVYLKAPNSSQCCFPMSDGPLFMHLLHMHLNDLLDSNAYSISLIKEEIELVSQELEFIRSFFGDAAEQGLYK
8. MDADESILIAKDTKLENLRILGELLISYSKDTMNIFKRFPNLQVLQFELKESWDYSTEQHWFPKLDCLTELETLCVGFKSSNTNHCGSSVATNRPWDFHF
9. ICFNEIGDILNFQIHDLVHDFCLIKARKENLFDRIRSSAPSDLLPRQITIDYDEEEEHFGLNFVMFDSNKKRHSGKHLYSLGINGDQLDDSVSDAFHLRH
10. IVPTELEVMHICYTNLKASTSAEVGRFIKKLLETSPDILREYIIQLQEHMLTVIPPSTLGARNIHVMMEFLLLILSDMPKDFIHHDKLFDLLAHVGTLTR
11. PEDVAETCILELVERNLIDVVRWKPDGSPKSC
12. ARDIEDLLRSLTAKRDYSLLRVLDLEGVYKPMLPEVLGKLLLL
13. CPAELLDIGKEIAEKCKGSPLAISLIGG
14. SLLPLNLRVLTLSMSELKEDPMPQLGKL
15. IWARVLDVAYEAKDVIDSIIVRDNGLLHLIFSLPITIKKIKLIKEEISALDENIPKDRGLIVVNSPKKPVERKSLSTDKIIVGFEEETNLILRKLTSGPA
16. SPMLNVLRKLTNLKKLGLTCHPKAVKEVTECISYLTKLQSLKLRSRDLFGQPSDLELSDMRGLTLSLNLYLLG
17. MEKRKDIEEANNSLVLFSALSKDIANVLIFLENEENQKALDKDQVEKLKLKMAFICTYVQLSYSDFEQFEDIMTRNRQEV
18. EPPLIMVNDSLLNEICMLNHLRYLRIRTQVKYLPFSFSNLWNLESLFVSNKGSILVLLPRILDLVKLRVLSVGACSFFD
19. QECGKLEEIPPSFGDIYSLKFIKIVKSPQLEDSALKIKKYAEDMRGGNELQILGQ
20. AMIQGFFSTRDRPSEGKEWLGRLLRAMHEAEDFIDKFNLREARGRREALHVVTRPAAELISKCKLWRDLSNLVSKMQELCZDQY

431

432 **Figura 7.** Twenty conserved motifs were identified in the AAC97933.1 protein (CNL,
433 *Mi-1* gene) conferring resistance to *Meloidogyne spp.;* ABE68835.1 protein (CNL, *CaMi*
434 gene) conferring resistance to *M. incognita*; CAD29729.1 protein (CNL, *Hero* gene)
435 conferring resistance to *G. pallida* and *G. rostochiensis*; AAF04603.1 protein (CNL,
436 *Gpa2* gene) conferring resistance to *G. pallida*; and the NBS-LRR proteins
437 evm.model.ctg15.1201 (XNL), evm.model.ctg9.85 (CNL), evm.model.ctg40.1964
438 (XNL), evm.model.ctg40.1963 (CNL), evm.model.ctg40.2149 (NL),
439 evm.model.ctg13.1155 (NL), evm.model.ctg40.2154 (NL), and evm.model.ctg40.2144
440 (N) from *P. guajava*.

441

Finally, the NBS-LRR proteins evm.model.ctg17.337 (CNL), evm.model.ctg17.339 (CN), evm.model.ctg17.297 (XNL), evm.model.ctg17.321 (CNL), evm.model.ctg15.1416 (CNL), evm.model.ctg15.1452 (XNL), evm.model.ctg15.1388 (XNL), evm.model.ctg15.544 (XNL), evm.model.ctg13.687.1 (CNL), evm.model.ctg40.637 (CN), evm.model.ctg6.91 (XNL) from *P. guajava*; *Arabidopsis thaliana* proteins AT3G14460 (CNL) and AT3G14470 (CNL); CAD12795.1 (NL, *Cre3* gene) and AAM94164.1 (NL, *Cre1* gene) conferring resistance to *H. avenae*, which formed the fourth highlighted clade, exhibited conserved motifs including P-loop, Kinase-2, Kinase-3, GLPL, and RNBS-D (Figure 8).

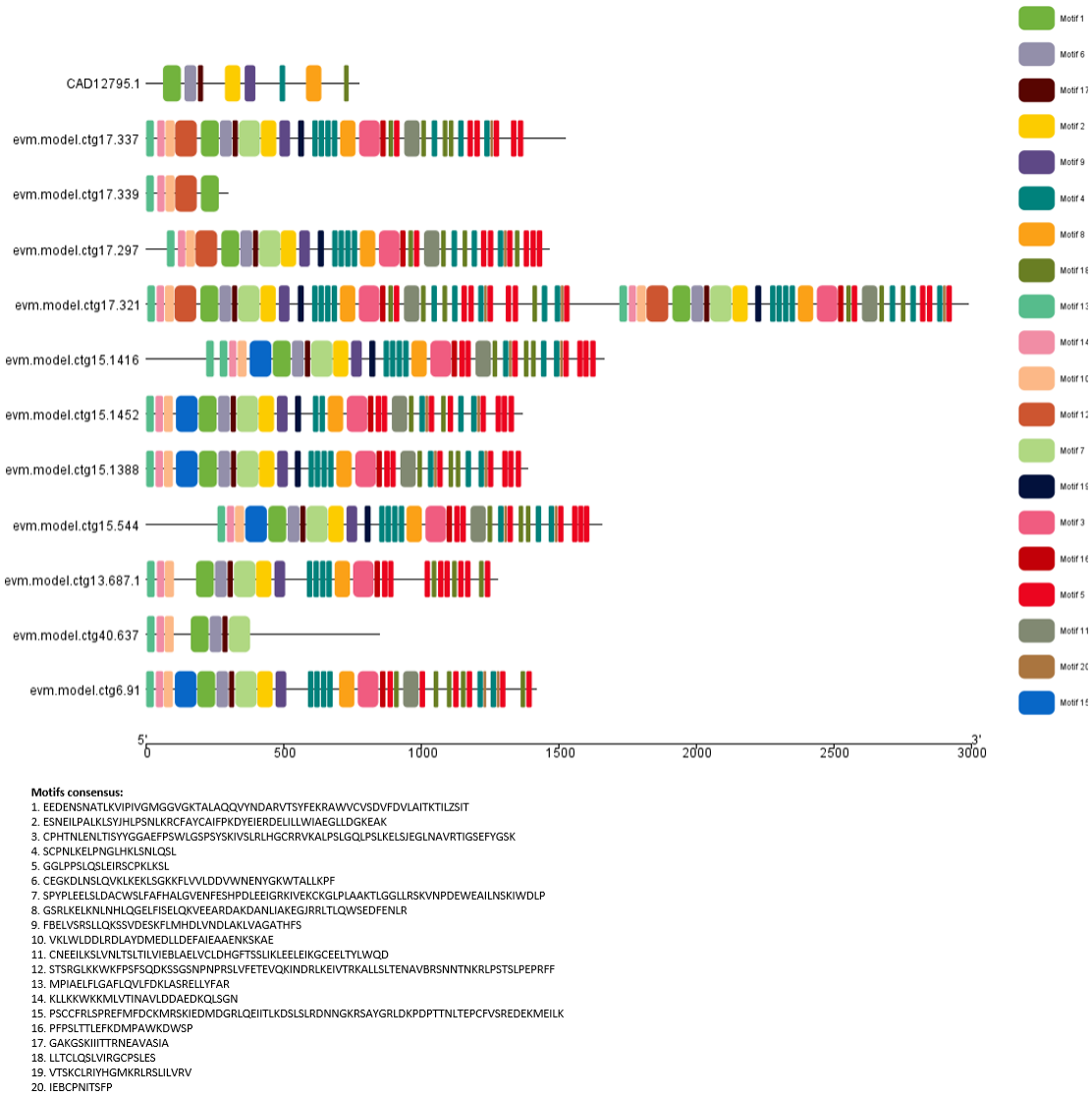


Figure 8. Twenty conserved motifs of the CAD12795.1 protein (NL, *Cre3* gene) and AAM94164.1 protein (NL, *Cre1* gene) conferring resistance to *H. avenae*, and the NBS-LRR proteins evm.model.ctg17.337 (CNL), evm.model.ctg17.339 (CN),

evm.model.ctg17.297 (XNL), evm.model.ctg17.321 (CNL), evm.model.ctg15.1416 (CNL), evm.model.ctg15.1452 (XNL), evm.model.ctg15.1388 (XNL), evm.model.ctg15.544 (XNL), evm.model.ctg13.687.1 (CNL), evm.model.ctg40.637 (CN), and evm.model.ctg6.91 (XNL) from *P. guajava*, along with *Arabidopsis thaliana* proteins AT3G14460 (CNL) and AT3G14470 (CNL).

The NBS domain was the most conserved region among *P. guajava*, *A. thaliana*, *O. europaea*, and functionally characterized nematode resistance (R) proteins, predominantly exhibiting conserved motifs such as Kinase-2, GLPL, MHDV, Kinase-3, RNBS-A, L5, L6, and P-loop (Supplementary Material 1 – Table 4).

3.4. Microsatellite marker analysis.

In summary, 25 microsatellites were found among the 279 NBS-LRR protein sequences of *P. guajava* analyzed (Figure 9; Supplementary Materials 2 and 6, Tables 1-7). Of these, 23 are simple SSRs; two sequences contain more than one SSR, with one classified as a compound SSR due to the proximity of motif regions being less than or equal to 100 nucleotides. Among the SSRs, nine (36%) are monomers, five (20%) are dimers, and eleven (44%) are trimers. Eight SSRs are repeated 6 times, five are repeated 10 times, and four are repeated 5 times. The most frequent repeat counts are six, ten, and five, respectively.

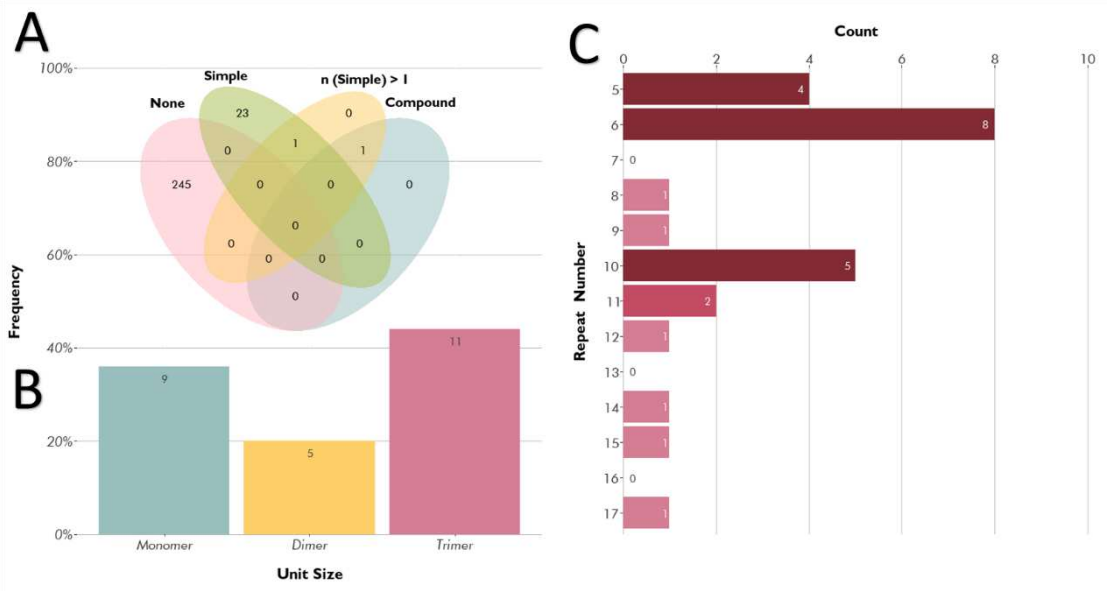


Figure 9. Microsatellite marker analysis in NBS-LRR proteins of *Psidium guajava*. **A.** Venn diagram illustrating the identification of 25 microsatellites (SSRs) among the 279 NBS-LRR protein sequences analyzed. (Legend: pink – none; green – simple; yellow – n (simple) > 1; blue – compound.) **B.** Distribution of SSR types: 9 monomers (blue), 5 dimers (yellow), and 11 trimers (pink). **C.** Frequency chart showing the number of SSR motif repeats.

Among the 25 microsatellites analyzed, 19 were detected within complete genes, predominantly belonging to the TNL subclass. Except for two proteins, these were classified into a gene family by InterProScan as members of the "Disease resistance protein, plants" family (IPR044974)". Noteworthy examples include a microsatellite identified in the protein evm.model.ctg13.1155, which is orthologous to the resistance proteins encoded by the *Mi-1* (CNL), *CaMi* (NL), *Hero* (CNL), and *Gpa2* (CNL) genes; and another microsatellite detected in the protein evm.model.ctg15.1452, orthologous to resistance proteins encoded by the *Cre3* (NL) and *Cre1* (NL) genes and the *Arabidopsis* resistance proteins AT3G14460 (CNL) and AT3G14470 (CNL). Additionally, a microsatellite was identified in the protein evm.model.ctg10.132, which occupies an isolated position on the second branch of the phylogeny.

3.5.Variant calling analysis

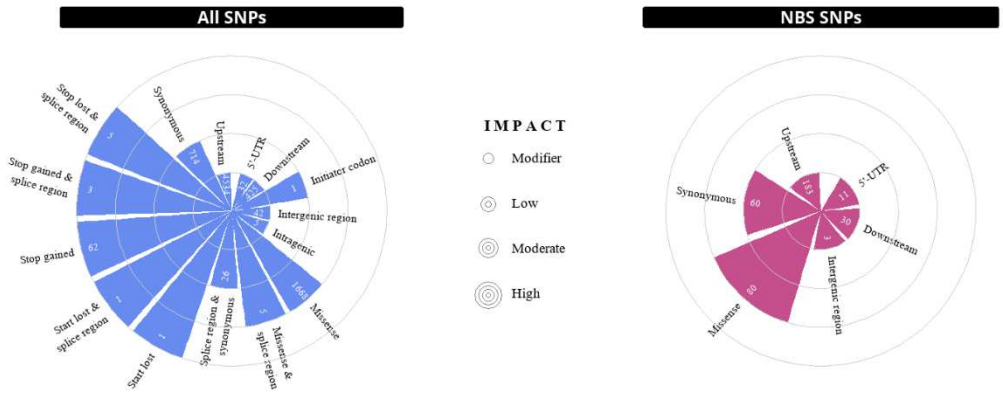


Figure 10. SNP distribution in *Psidium guajava*. Most variants were modifiers in non-

coding regions, while NBS-LRR genes showed enrichment in synonymous and missense SNPs

The variant calling analysis identified a diverse distribution of SNPs across the *P. guajava* genome, both in the full dataset and in the subset corresponding to NBS-LRR genes. For all SNPs, the majority of variants were classified as modifiers, predominantly located in intergenic, intragenic, and upstream regions, with a smaller fraction of coding variants such as synonymous and missense substitutions. The moderate- and high-impact categories were mainly represented by missense, stop-gained, and splice-region variants, which are more likely to alter protein function. In contrast, when focusing on NBS-LRR genes, a distinct profile was observed: the variants were concentrated in missense and synonymous categories, suggesting that these resistance-related genes are hotspots of both functional and potentially adaptive variation. The predominance of moderate-impact (missense) SNPs in NBS-LRR genes may reflect selective pressure maintaining functional diversity within this gene family, consistent with their role in pathogen recognition and defense response in plants.

4. Discussion

Chemical treatments remain the predominant strategy for disease control in crop production (Chen et al. 2007). However, the indiscriminate use of these compounds has contributed to environmental degradation and the emergence of more resistant pathogen strains. In contrast, the development of disease-resistant varieties remains underexplored in guava cultivation. The genome-wide identification of NBS-LRR genes presents a promising avenue for the discovery of functional resistance (R) genes through high-throughput screening approaches. NBS-LRR proteins play a central role in plant immune responses (Sharma et al. 2017; Chang et al. 2020) mediating defense signaling against a wide array of pathogens (Meyers et al. 2003a; Zhou et al. 2004; Jupe et al. 2012; Lozano et al. 2012; Wei et al. 2013; Gu et al. 2015; Lozano et al. 2015; Christie et al. 2016; Wang et al. 2017; Zhang et al. 2020a; Chang et al. 2020; Inturrisi et al. 2020). Furthermore, the NBS-LRR gene family constitutes the largest group of characterized resistance genes (Hammond-Kosack and Jones 1997; Sagi et al. 2017) and is considered one of the most significant R gene families (Jorgensen and Emerson 2009; Saha et al. 2013; Hwang et al. 2014; Shi et al. 2018).

In this study, NBS-LRR genes were identified and characterized for the first time in the guava genome, which spans 443.76 Mb and contains 25,601 protein-coding genes. These genes represent approximately 1.09% of the genome's protein-coding sequences. Previous studies have reported an expected frequency of 0.6% to 1.76% for NBS sequences across several plant genomes, placing the frequency observed in guava within this range (Porter et al. 2009; Lozano et al. 2015). However, many plant species exhibit NBS gene frequencies outside this range, such as approximately 0.5% in *Arabidopsis thaliana* (Meyers et al. 2003a), 0.58% in *Glycine max* (soybean) (Kang et al. 2012), 2.2% in *Coffea canephora* (coffee) (Denoeud 2014), 0.9% in *Manihot esculenta* (cassava) (Lozano et al. 2015), 0.43% in *Cicer arietinum* (chickpea) (Sagi et al. 2017), 0.67% in *Helianthus annuus* (sunflower) (Neupane et al. 2018), 0.58% to 0.75% in *Brassica* species (Inturrisi et al. 2020), and 0.5% in *Olea europaea* (olive) (Bettaieb and Bouktila 2020). Such variations may result from differences in genome annotation quality, identification criteria, prediction tools, and filtering thresholds (Inturrisi et al. 2020). Additionally, genome size (Bettaieb and Bouktila 2020), the dynamic evolutionary history of NBS-LRR genes (Gu et al. 2015), and the necessity for adaptation against diverse pathogens (Kapos et al. 2019) may contribute to the expansion or contraction of these gene subfamilies. Genetic mechanisms such as whole-genome duplications, tandem gene duplications, and segmental duplications likely contribute to these patterns (Sharma et al. 2017).

The analysis of conserved domains allowed the classification of guava NBS-LRR genes into subclasses, with a higher proportion of TIR (TNL and TN) subclasses (54.1%) compared to non-TIR subclasses (CNL, CN, XNL, XN, RNL, RN, NL, and N) (45.87%). This distribution is comparable to that observed in *Eucalyptus grandis*, with 43.4% TIR and 16.1% complete non-TIR genes (Christie et al. 2016), *Malus* × *domestica*, with 69.92% TIR and 30.08% non-TIR genes (Arya et al. 2014), and *Arabidopsis thaliana*, with 70.05% TIR and 29.95% non-TIR genes (Meyers et al. 2003a). The predominance of TIR-type genes suggests their significant role in resistance mechanisms against pathogens common to these species (Leister 2004) and indicates a more recent TIR expansion, potentially driven by tandem duplications events (Christie et al. 2016).

However, this pattern is reversed in some woody and herbaceous species, such as *Gossypium hirsutum* (Shi et al. 2018) and *Olea europaea* (Bettaieb and Bouktila 2020). Some studies propose that TIR-type genes preceded the emergence of non-TIR genes, suggesting that dicot NBS-LRR genes may have evolved from ancestral TIR forms (Sharma et al. 2017; Bettaieb and Bouktila 2020). In contrast, other evidence indicates that CNL genes represent the most ancient and diverse lineage within the NBS-LRR family (Wang et al. 2020). The TIR subclass is predominantly associated with dicot species and is rarely identified in monocots and magnoliids (Tarr and Alexander 2009; Bettaieb and Bouktila 2020). Additionally, Lozano et al. (2015) reported that most CNL genes in grasses lack homologs in dicots, highlighting significant divergent evolution of NBS-LRR genes between monocots and dicots, and furthermore, the rapid evolution of these genes is also evident among different cultivars of the same species under varying biotic stress conditions.

Approximately 60.57% of the NBS-LRR proteins identified in guava are classified as non-canonical, including truncated NBS-LRRs, NBS-LRRs with atypical domains, and those with atypical assemblies. Truncated NBS-LRR genes have been reported in several plant species, likely arising from partial deletions of redundant NBS-LRR genes (Wang et al. 2020). Such truncations may also result from pseudogenization driven by the rapid evolution of the NBS-LRR gene family (Lozano et al. 2015). Truncated genes have been shown to function cooperatively with full-length NBS-LRRs. For instance, in *Arabidopsis thaliana*, the tandem NBS-LRR genes RPP2A (TIR-NBS, truncated) and RPP2B (full-length) are both required to confer resistance to *Peronospora parasitica* (Sinapidou et al. 2004; Sagi et al. 2017).

In addition to the CC, TIR, NBS, LRR, and RPW8 domains of powdery mildew resistance proteins (PF05659), atypical structural domains were detected in 18 predicted NBS-LRR proteins (6.45%). Similar findings have been reported in other species, such as *Olea europaea* (Bettaieb and Bouktila 2020). The identification of such domains is essential for refining the classification of proteins in this gene family and for elucidating their roles in signal transduction during plant resistance to biotic stress. The fusion of NBS-LRR domains with additional domains is believed to enhance pathogen effector detection and recognition (Sarris et al. 2015), providing functional versatility and plasticity to this gene family (Bettaieb and Bouktila 2020). Genes with atypical assemblies have also been identified in other genomes, such as Brassica juncea, where they may play roles in plant immunity. Further investigations are needed to clarify their involvement in plant resistance mechanisms (Inturrisi et al. 2020).

The chemical and structural characteristics of NBS-LRR genes and their encoded proteins in guava exhibited substantial variation. Similar patterns have been reported in studies of NBS genes from cucumber (*Cucumis sativus* L.) (Zhang et al. 2022), banana (*Musa spp.*) (Chelliah et al. 2023), grapevine (*Vitis vinifera*) (Goyal et al. 2020), cabbage (*Brassica pekinensis*) (Liu et al. 2021), and mungbean (*Vigna radiata* L.) (Purwar et al. 2023). Subcellular localization predictions indicated that NBS-LRR proteins could be distributed across multiple organelles, with most proteins located in the nucleus (45.16%), chloroplast (22.94%), and cytoplasm (22.22%). This distribution is consistent with findings in sunflower (Neupane et al. 2018), kiwifruit (*Actinidia chinensis*) (Wang et al. 2020), and cucumber (Zhang et al. 2022). The predominant nuclear localization reflects the role of these proteins in transcriptional regulation and pathogen recognition, while their functional versatility may be influenced by their specific chemical and structural attributes (Zhang et al. 2022).

606 Chromosomal mapping revealed a higher density of genes on chromosomes 3, 5,
607 and 8, which harbor 51 TIR genes (33.77% of guava TIR genes), 23 TIR genes (15.23%),
608 and 36 TIR genes (23.84%), respectively, accounting for 72.84% of the total TIR genes
609 in the species. Clusters of this subclass were also observed on these chromosomes.
610 Previous studies have shown that NBS-LRR genes typically form tightly linked clusters
611 with high sequence homology (Lozano et al. 2015; Neupane et al. 2018; Wang et al.
612 2020), likely resulting from duplications, including whole-genome and tandem
613 duplications (Wang et al. 2020). These findings suggest that gene duplications on
614 chromosomes 3, 5, and 8 are more prevalent in guava compared to other chromosomes.
615 In addition to duplications, evolutionary events affecting NBS-LRR genes, such as
616 unequal crossing-over, insertion/deletion, and gene conversion, may increase mutation
617 rates and gene density on specific chromosomes, as observed in radish (*Raphanus sativus*
618 L.) (Ma et al. 2021) and cucumber. Such genomic dynamics are likely to enhance the
619 plant's ability to defend against diverse pathogens in nature (Zhang et al. 2022).

620 The phylogenetic analysis revealed a clear evolutionary divergence between the
621 TIR and non-TIR subclasses, in agreement with previous studies (Meyers et al. 2003a;
622 Lozano et al. 2015; Sagi et al. 2017). Notably, the RNL and RN subclasses, except for
623 the RNL protein evm.model.ctg9.2026, formed a distinct subclade within the CNL clade,
624 suggesting functional divergence of the RPW8 domain from typical CNL genes (Collier
625 et al. 2011; Sagi et al. 2017). Clades of complete and partial NBS-LRR proteins, located
626 on the same chromosomes, were observed, indicating potential recent gene duplications
627 to expand the resistance gene repertoire in guava, similar to findings in *Eucalyptus*
628 *grandis* (Christie et al. 2016). However, clustering analysis is needed to confirm these
629 patterns. Some phylogenetic clusters included exclusively *P. guajava* proteins, while
630 others comprised proteins from *P. guajava*, *Arabidopsis*, and other resistance genes,
631 suggesting shared evolutionary origins and sequence homology. Furthermore, both
632 phylogenetic relationships and conserved motif analyses corroborated the classification
633 into TNL, TN, CNL, CN, RNL, RN, XNL, XN, NL, and N subclasses.

Notably, 29 guava NBS-LRR proteins showed orthology to nematode resistance proteins. Among them, the TNL protein evm.model.ctg40.285 clustered closely with the TNL protein R AAP44390.1, encoded by the *Gro1-4* gene in *S. tuberosum*, conferring resistance to *G. rostochiensis*. Both proteins share similar sequence lengths (1066 and 1136 amino acids, respectively) and conserved motifs such as P-loop and kinase-3. This suggests that evm.model.ctg40.285 may be encoded by a resistance gene locus in guava, similar to the *Gro1-4* locus in potatoes, where all members belong to the TNL subclass.

Furthermore, clusters of highly homologous resistance genes may provide defense against multiple pathogens (Van Der Vossen et al. 2000). Eight guava NBS-LRR proteins (2 XNL, 2 CNL, 3 NL, and 1 N) were orthologous to nematode resistance proteins from other species, including AAC97933.1 (CNL) (*Mi-1*) from *S. lycopersicum*, ABE68835.1 (CNL) (*CaMi*) from *C. annuum*, CAD29729.1 (CNL) (*Hero*) from *S. lycopersicum*, and AAF04603.1 (CNL) (*Gpa2*), conferring resistance to several nematodes. These proteins shared 20 conserved motifs, including RNBS-A, kinase-2, kinase-3, L5, L6, P-loop, and C1. Additionally, eleven NBS-LRR proteins (4 CNL, 2 CN, and 5 XNL) showed orthology to the resistance proteins R CAD12795.1 (*Cre3*) and AAM94164.1 (*Cre1*) from *Triticum sp.,* conferring resistance to *H. avenuae*, with shared motifs such as P-loop, RNBS-A, kinase-2, kinase-3, GLPL, and MHDV.

Finally, nine guava NBS-LRR proteins (4 CNL and 5 XNL) were found to be orthologous to ADP37705 (TNL), encoded by *Rpm2* in *P. sogdiana*, conferring resistance to *M. incognita*. These proteins exhibited 20 conserved motifs, including P-loop, kinase-2, kinase-3, GLPL, and RNBS-D. Homologous genes can be classified as orthologs - genes diverging through speciation that generally retain similar functions - and paralogs - genes diverging through duplication that often acquire novel functions (Fitch 1970; Tatusov et al. 1997). This distinction facilitates the functional characterization of new resistance genes (R genes) in *P. guajava* when analyzed alongside well-characterized NBS-LRR genes.

The conserved motif analysis of NBS-LRR proteins, as highlighted in the phylogenetic analysis, revealed significant amino acid sequence similarity. The NBS domain, a region of up to 300 amino acids, consists of eight conserved motifs: P-loop (phosphate-binding loop), GLPL (Gly-Leu-Pro-Leu, also known as kinase 3), RNBS-A, RNBS-B, RNBS-C, RNBS-D, kinase 2, and MHDV (Met-His-Asp-Val). Although the

degree of conservation of these motifs varies across plant species (Meyers et al. 2003a; Zhang et al. 2022), the NBS domain remains the most conserved region among domains encoded by disease resistance genes in plants (Arya et al. 2014). A known distinction exists in the conserved motifs between TNL and CNL classes, with different patterns depending on whether they belong to the TIR or non-TIR groups (Arya et al. 2014).

Our analyses corroborate these findings, detecting conserved motifs such as GLPL, kinase 2, kinase, P-loop, RNBS-A, and RNBS-B within the NBS domains of non-TIR NBS-LRR genes. Conversely, the P-loop motif was consistently conserved in the NBS domains of TIR NBS-LRR genes across *P. guajava*, *A. thaliana* (Meyers et al. 2003a), and *O. europaea* (Bettaieb and Bouktila 2020). However, a notable difference was observed in the MHDV motif. In *P. guajava* and *O. europaea*, valine (V) is replaced by leucine (L), resulting in the motif MHDL, a variation also reported in *M. domestica* (Arya et al. 2014). In contrast, the LRR and CC domains of *P. guajava* NBS-LRR genes showed lower motif conservation. In the LRR domain, only L5 and L6 motifs were observed, consistent with findings in *O. europaea*.

Microsatellites located within functionally relevant genes, such as those from the NBS-LRR family, may serve as indicators of genetic variability linked to pathogen resistance. In this study, 25 SSRs were identified among 279 NBS-LRR protein sequences from *Psidium guajava*. The majority of SSRs were simple type repeats, with a predominance of trinucleotide motifs (44%), followed by mononucleotide (36%) and dinucleotide (20%) motifs. The most common repeat numbers were 6, 10, and 5, suggesting the presence of recurring and potentially functional patterns in these sequences.

Nineteen of the 25 SSRs were located within full-length genes, predominantly belonging to the TNL subclass, which is frequently associated with resistance against biotrophic pathogens in dicots(Meyers et al. 2003b; Sela et al. 2012). Most of these SSR-containing genes were annotated as belonging to the "Disease resistance protein, plants" family (IPR044974), reinforcing the possible role of these repetitive elements in regulating immune responses in *P. guajava*.

Notably, two SSRs were found in genes orthologous to well-characterized nematode resistance genes. One SSR was located in the protein evm.model.ctg13.1155,

orthologous to *Mi-1*, *CaMi*, *Hero*, and *Gpa2*, all known to confer resistance against *Meloidogyne* and *Globodera* species (Milligan et al. 1998; Van Der Vossen et al. 2000; Ernst et al. 2002). The second SSR was found in evm.model.ctg15.1452, orthologous to *Cre1* and *Cre3*, which confer resistance to Heterodera in *Beta vulgaris* (De Majnik et al. 2003).

The presence of SSRs in orthologs of well-established resistance loci highlights their potential as functional markers. These SSRs could be valuable in marker-assisted selection (MAS) for guava breeding programs (Arabzai, Muhammad Gul; Gul 2021), particularly considering the impact of plant-parasitic nematodes on yield (Sikandar et al. 2020). Moreover, these markers can be applied in intraspecific diversity studies, QTL mapping for resistance traits, and even in gene editing strategies aimed at introducing or enhancing resistance (Collard et al. 2005; Kalia et al. 2011).

In summary, the SSRs identified within NBS-LRR genes of *P. guajava* represent a valuable source of molecular markers for genetic studies and breeding. Their localization within orthologs of classic nematode resistance genes highlights promising targets for future functional validation.

The genome-wide variant analysis in *Psidium guajava*—showing mostly modifier SNPs in intergenic/upstream regions and a smaller fraction of coding (synonymous and missense) changes—matches patterns reported for Myrtaceae where non-coding variation predominates but functionally relevant variants occur at resistance loci (Soewarto et al. 2019; Fernandes Santos et al. 2022). Predicted moderate- and high-impact effects (missense, stop-gained, splice-region) are consistent with the SnpEff-based annotations and with previous Myrtaceae studies that identify NBS-LRR clusters as hotspots of nonsynonymous variation linked to pathogen response (Fernandes Santos et al. 2022). Focusing on NBS-LRR genes, the observed enrichment for missense and synonymous SNPs suggests balancing/ diversifying selection maintaining recognition diversity — a pattern described for *Eucalyptus* and other Myrtaceae and supported by the comparative and transcriptome variant analyses in the attached papers (Soewarto et al. 2019; Fernandes Santos et al. 2022).

5. Conclusions

This study provides novel insights into the NBS-LRR gene family, the key group involved in plant disease resistance, within *Psidium guajava*, a crop of significant global socioeconomic importance. Additionally, orthologs of nematode resistance genes were identified, providing foundational information to enhance our understanding of *P. guajava*-nematode interactions and potentially identifying candidate resistance genes. Microsatellite analysis revealed SSRs in NBS-LRR proteins, including two genes orthologous to well-characterized nematode resistance genes, which may serve as promising molecular markers for future breeding programs. Variant calling analysis indicate that NBS-LRR genes are hotspots of functional variation, possibly under selective pressure, reinforcing their role in maintaining diversity and defense response in *P. guajava*. These resources are valuable for advancing the functional exploration of these genes and ultimately uncovering their roles in defending against various pathogens.

Acknowledgments

We would like to thank all colleagues at the Biometrics and Genetics and Plant Improvement laboratories from the Universidade Federal do Espírito Santo, as well as the funding agencies Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil); and Fundação de Amparo à Pesquisa do Espírito Santo (FAPES, Vitória—ES, Brazil).

Author contributions

Layra de Medeiros Cardozo: acquisition of data; analysis and interpretation of data; drafting of the article; Gilza Barcelos de Souza: acquisition of data; analysis and interpretation of data; drafting of the article; Iana Pedro da Silva Quadros: acquisition and analysis of data; Guadalupe del Rosario Quispe Saji: acquisition and analysis of data; Giovana Pinto Pires: analysis and interpretation of data; Pedro Henrique Dias dos Santos: analysis and interpretation of data; Adésio Ferreira: conception and design of the study; interpretation of data; Marcia Flores da Silva Ferreira: conception and design of the study; interpretation of data; critical revision for important intellectual content. All authors read and approved the final manuscript.

Funding: This work was supported by the Conselho Nacional de Pesquisa (CNPq, Brazil) Finance Code 200329/2021-9; the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) - Finance Code 001; and the Fundação de Amparo à Pesquisa do Espírito Santo (FAPES, Vitória—ES, Brazil).

Data Archiving Statement

The nucleotide sequences of the NBS-LRR genes identified in *Psidium guajava* in this study, as well as the corresponding amino acid sequences of their encoded proteins, are being submitted to the NCBI GenBank database. Accession numbers will be provided once the submission process is complete.

Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

Alizadeh-Moghaddam G, Nasr-Esfahani M, Nasr-Esfahani A, Sedaghatfar E, Rahanandeh H, Yazdi MJ (2024) Comparative genetic defence analysis using microsatellite markers and anatomical resistance of potato cultivars to early blight. Physiol Mol Plant Pathol 133:102374. https://doi.org/10.1016/J.PMPP.2024.102374

Andrews S (2010) Andrews, S. (2010). FastQC: a quality control tool... - Google Scholar. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed 20 May 2020

Arabzai, Muhammad Gul; Gul H (2021) Application techniques of molecular marker and achievement of marker assisted selection (MAS) in three major crops rice, wheat and maize. Int J Res Appl Sci Biotechnol 8:82–93. https://doi.org/https://doi.org/10.31033/ijrasb.8.1.10

Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., De Castro, E., ... & Stockinger H (2012) ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 40:W597–W603. https://doi.org/10.1093/nar/gks400

Arya P, Kumar G, Acharya V, Singh AK (2014) 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:e107987. https://doi.org/10.1371/JOURNAL.PONE.0107987

Bakker E, Dees R, Bakker J, Goverse A (2006) Mechanisms involved in plant resistance to nematodes. Springer 314–334. https://doi.org/10.1007/0-387-23266-4_14

Bashir MH, Afzal M, Shakeel Nawaz M, Khan AZ, Aziz S, Ullah F, Amjad I (2022) Characterization and advancement of microsatellite (SSR) markers for various stresses in wheat. https://doi.org/10.47278/journal.ijab/2022.018

Beier S, Thiel T, Münch T, Scholz U, Bioinformatics MM-, 2017 U (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics. https://doi.org/10.1093/bioinformatics/btx198

Bentham AR, de la Concepcion JC, Mukhi N, Zdrzałek R, Draeger M, Gorenkin D, Hughes RK, Banfield MJ (2020) A molecular roadmap to the plant immune system. J Biol Chem 295:14916–14935. https://doi.org/10.1074/JBC.REV120.010852

Bettaieb I, Bouktila D (2020) Genome-wide analysis of NBS-encoding resistance genes in the Mediterranean olive tree (Olea europaea subsp. europaea var. europaea): insights into their molecular diversity, evolution and function. Tree Genet Genomes 16:23. https://doi.org/10.1007/S11295-020-1415-9

Bigeard J, Colcombet J, Hirt H (2015) Signaling mechanisms in pattern-triggered immunity (PTI). Mol Plant 8:521–539. https://doi.org/10.1016/J.MOLP.2014.12.022

Bolger A, Lohse M, Bioinformatics BU-, 2014 U (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120

Chang W, Li H, Chen H, Qiao F, Zeng H (2020) NBS-LRR gene family in banana (Musa acuminata): genome-wide identification and responses to Fusarium oxysporum f. sp. cubense race 1 and tropical race 4. Eur J Plant Pathol 157:549–563. https://doi.org/10.1007/S10658-020-02016-7

Chelliah A, Arumugam C, Suthanthiram B, Raman T, Subbaraya U (2023) Genome-wide identification, characterization, and evolutionary analysis of NBS genes and their association with disease resistance in Musa spp. Funct Integr Genomics 23:7. https://doi.org/10.1007/S10142-022-00925-W

Chen R, Li H, Zhang L, Zhang J, Xiao J, Ye Z (2007) CaMi, a root-knot nematode resistance gene from hot pepper (Capsium annuum L.) confers nematode resistance

827 in tomato. Plant Cell Rep 26:895–905. https://doi.org/10.1007/S00299-007-0304-0
828 Chen S, Zhou Y, Chen Y, Gu J (2018) fastp: an ultra-fast all-in-one FASTQ
829 preprocessor. Bioinformatics 34:i884.
830 https://doi.org/10.1093/BIOINFORMATICS/BTY560
831 Christie N, Tobias PA, Naidoo S, Külheim C (2016) The Eucalyptus grandis NBS-LRR
832 gene family: Physical clustering and expression hotspots. Front Plant Sci 6:1238.
833 https://doi.org/10.3389/FPLS.2015.01238/FULL
834 Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden
835 DM (2012) A program for annotating and predicting the effects of single
836 nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila
837 melanogaster strain w1118; iso-2; Taylor Fr Cingolani, A Platts, LL Wang, M
838 Coon, T Nguyen, L Wang, SJ Land, X Lu, DM Rudenfly, 2012•Taylor Fr 6:80–92.
839 https://doi.org/10.4161/FLY.19695
840 Claverie M, Dirlewanger E, Bosselut N, van Ghelder C, Voisin R, Kleinhentz M,
841 Lafargue B, Abad P, Rosso MN, Chalhoub B, Esmenjaud D (2011) The ma gene
842 for complete-spectrum resistance to meloidogyne species in prunus is a TNL with
843 a huge repeated c-terminal post-LRR region. Plant Physiol 156:779–792.
844 https://doi.org/10.1104/PP.111.176230
845 Cleary JG, Braithwaite R, Gaastra K, Hilbush BS, Inglis S, Irvine SA, Jackson A, Littin
846 R, Rathod M, Ware D, Zook JM, Trigg L, De La Vega FMM (2015) Comparing
847 Variant Call Files for Performance Benchmarking of Next-Generation Sequencing
848 Variant Calling Pipelines. bioRxiv 023754. https://doi.org/10.1101/023754
849 Coll N, Epple P, Dangl J (2011) Programmed cell death in the plant immune system.
850 Cell Death Differ 18:1247–1256. https://doi.org/10.1038/cdd.2011.37
851 Collard BCY, Grams RA, Bovill WD, Percy CD, Jolley R, Lehmensiek A, Wildermuth
852 G, Sutherland MW (2005) Development of molecular markers for crown rot
853 resistance in wheat: mapping of QTLs for seedling resistance in a '2-
854 49'×'Janz'population. Plant Breed 124:532–537. https://doi.org/10.1111/J.1439-
855 0523.2005.01163.X
856 Collier SM, Hamel LP, Moffett P (2011) Cell death mediated by the N-terminal
857 domains of a unique and highly conserved class of NB-LRR protein. Mol Plant-
858 Microbe Interact 24:918–931. https://doi.org/10.1094/MPMI-03-11-0050
859 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A,
860 Keane T, McCarthy SA, Davies RM (2021) Twelve years of SAMtools and

BCFtools. Gigascience 10:giab008. https://doi.org/10.1093/GIGASCIENCE/GIAB008

De Majnik J, Ogbonnaya FC, Moullet O, Lagudah ES (2003) The Cre1 and Cre3 Nematode Resistance Genes Are Located at Homeologous Loci in the Wheat Genome. Mol Plant-Microbe Interact 16:1129–1134. https://doi.org/10.1094/MPMI.2003.16.12.1129

Delorenzi M, Speed T (2002) An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. Bioinformatics 18:617–625. https://doi.org/10.1093/bioinformatics/18.4.617

Denoeud F et al. (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science (80-) 345:1181–1184. https://doi.org/10.1126/science.1255274

Djian-Caporalino C, Pijarowski L, Fazari A, Samson M, Gaveau L, O'Byrne C, Lefebvre V, Caranta C, Palloix A, Abad P (2001) High-resolution genetic mapping of the pepper (Capsicum annuum L.) resistance loci Me3 and Me4 conferring heat-stable resistance to root-knot nematodes. Springer 103:592–600. https://doi.org/10.1007/PL00002914

Doyle D& (1990) Isolation of plant DNA from fresh tissue. Focus (Madison) 12:13–15

Ernst K, Kumar A, Kriseleit D, Kloos D-U, Phillips MS, Ganal MW (2002) The broad-spectrum potato cyst nematode resistance gene (Hero) from tomato is the only member of a large gene family of NBS-LRR genes with an unusual amino. Plant J 31:127–136. https://doi.org/10.1046/j.1365-313X.2002.01341.x

Ewels P, Magnusson M, Lundin S, Käller M (2016) MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32:3047. https://doi.org/10.1093/BIOINFORMATICS/BTW354

Feng C, Feng C, Lin X, Liu S, … YL (2021) A chromosome-level genome assembly provides insights into ascorbic acid accumulation and fruit softening in guava (Psidium guajava). Plant Biotechnol J 19:717–730. https://doi.org/10.1111/pbi.13498

Fernandes Santos CA, da Costa SR, Boiteux LS, Grattapaglia D, Silva-Junior OB (2022) Genetic associations with resistance to Meloidogyne enterolobii in guava (Psidium sp.) using cross-genera SNPs and comparative genomics to Eucalyptus highlight evolutionary conservation across the Myrtaceae. PLoS One 17:1–20. https://doi.org/10.1371/journal.pone.0273959

895 Finn R, Bateman A, Clements J, Research PC (2014) Pfam: the protein families
896 database. Nucleic Acids Res 42:D222–D230. https://doi.org/10.1093/nar/gkt1223
897 Fitch WM (1970) Distinguishing homologous from analogous proteins. Syst Zool
898 19:99–113. https://doi.org/10.2307/2412448
899 Gouveia BC, Calil IP, Machado JPB, Santos AA, Fontes EPB (2017) Immune receptors
900 and co-receptors in antiviral innate immunity in plants. Front Microbiol 7:2139.
901 https://doi.org/10.3389/FMICB.2016.02139/FULL
902 Goyal N, Bhatia G, Sharma S, Garewal N, Upadhyay A (2020) Genome-wide
903 characterization revealed role of NBS-LRR genes during powdery mildew
904 infection in Vitis vinifera. Genomics 112:312–322.
905 https://doi.org/10.1016/j.ygeno.2019.02.011
906 Grossi LL, Fernandes M, Silva MA, de Oliveira Bernardes C, Tuler AC, dos Santos
907 PHD, Ferreira A, da Silva Ferreira MF (2021) DArTseq-derived SNPs for the
908 genus Psidium reveal the high diversity of native species. Tree Genet Genomes
909 17:1–13. https://doi.org/10.1007/S11295-021-01505-Y/FIGURES/5
910 Gu L, Si W, Zhao L, Yang S, Zhang X (2015) Dynamic evolution of NBS–LRR genes
911 in bread wheat and its progenitors. Mol Genet Genomics 290:727–738.
912 https://doi.org/10.1007/S00438-014-0948-8
913 Hammond-Kosack KE, Jones JDG (1997) Plant disease resistance genes. Annu Rev
914 Plant Biol 48:575–607. https://doi.org/10.1146/ANNUREV.ARPLANT.48.1.575
915 Hermenegildo P da S, Santos SA, Guimarães LMS, Pereira IC, Vidigal PMP, Badel JL,
916 Alfenas-Zerbini P, Mafia RG, Ferreira MAS V., Alfenas AC (2019) Draft
917 Genome Sequence of Erwinia psidii, Causal Agent of Bacterial Blight of Guava (
918 Psidium guava) and Dieback of Eucalypt (Eucalyptus spp.) . Microbiol Resour
919 Announc 8. https://doi.org/10.1128/MRA.01598-18
920 Hogenhout SA, Van Der Hoorn RAL, Terauchi R, Kamoun S (2009) Emerging
921 concepts in effector biology of plant-associated organisms. Mol Plant-Microbe
922 Interact 22:115–122. https://doi.org/10.1094/MPMI-22-2-0115
923 Horton P, Park K, Obayashi T, … NF (2007) WoLF PSORT: protein localization
924 predictor. Nucleic Acids Res 35:W585–W587. https://doi.org/10.1093/nar/gkm259
925 Hwang IS, Choi DS, Kim H, Kim DS, Hwang BK (2014) Pathogenesis-related protein
926 4b interacts with leucine-rich repeat protein 1 to suppress PR 4b-triggered cell
927 death and defense response in pepper. Plant J 77:521–533.
928 https://doi.org/10.1111/tpj.12400

929 IBGE (2020) Produção agropecuária. https://www.ibge.gov.br/explica/producao-
930 agropecuaria/goiaba/br. Accessed 24 Mar 2023
931 Institute B (2019) Picard toolkit. https://broadinstitute.github.io/picard/
932 Inturrisi F, Bayer PE, Yang H, Tirnaz S, Edwards D, Batley J (2020) Genome-wide
933 identification and comparative analysis of resistance genes in Brassica juncea. Mol
934 Breed 40. https://doi.org/10.1007/s11032-020-01159-z
935 Jiangtao C, Yingzhen K, … WQ-YC (2015) MapGene2Chrom, a tool to draw gene
936 physical map based on Perl and SVG languages. Yi Chuan= Hered 37:91–97.
937 https://doi.org/10.16288/j.yczz.2015.01.013
938 Jones J, Dangl J (2006) The plant immune system. Nature 444:323–329.
939 https://doi.org/10.1038/nature05286
940 Jorgensen T, Emerson B (2009) RPW8 and resistance to powdery mildew pathogens in
941 natural populations of Arabidopsis lyrata. New Phytol 182:984–993.
942 https://doi.org/10.1111/j.1469-8137.2009.02787.x
943 Jupe F, Pritchard L, Etherington GJ, MacKenzie K, Cock PJA, Wright F, Sharma SK,
944 Bolser D, Bryan GJ, Jones JDG, Hein I (2012) Identification and localisation of the
945 NB-LRR gene family within the potato genome. BMC Genomics 13:1–14.
946 https://doi.org/10.1186/1471-2164-13-75
947 Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an
948 overview of the recent progress in plants. Springer 177:309–334.
949 https://doi.org/10.1007/S10681-010-0286-9
950 Kang YJ, Kim KH, Shim S, Yoon MY, Sun S, Kim MY, Van K, Lee SH (2012)
951 Genome-wide mapping of NBS-LRR genes and their association with disease
952 resistance in soybean. BMC Plant Biol 12:1–13. https://doi.org/10.1186/1471-
953 2229-12-139
954 Kapos P, Devendrakumar KT, Li X (2019) Plant NLRs: From discovery to application.
955 Plant Sci 279:3–18. https://doi.org/10.1016/J.PLANTSCI.2018.03.010
956 Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid
957 multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res
958 30:3059–3066. https://doi.org/10.1093/nar/gkf436
959 Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K,
960 Jaccoud D, Hopper C, Aschenbrenner-Kilian M, Evers M, Peng K, Cayla C, Hok
961 P, Uszynski G (2012) Diversity arrays technology: a generic genome profiling
962 technology on open platforms. Springer 888:67–89. https://doi.org/10.1007/978-1-

963 61779-870-2_5
964 Knaus BJ, Grünwald NJ (2017) vcfr: a package to manipulate and visualize variant call
965 format data in R. Mol Ecol Resour 17:44–53. https://doi.org/10.1111/1755-
966 0998.12549
967 Leister D (2004) Tandem and segmental gene duplication and recombination in the
968 evolution of plant disease resistance genes. Trends Genet 20:116–122.
969 https://doi.org/10.1016/j.tig.2004.01.007
970 Letunic I, Bork P (2021) Interactive Tree Of Life (iTOL) v5: an online tool for
971 phylogenetic tree display and annotation. Nucleic Acids Res 49:W293–W296.
972 https://doi.org/10.1093/nar/gkab301
973 Li H (2011) A statistical framework for SNP calling, mutation discovery, association
974 mapping and population genetical parameter estimation from sequencing data.
975 Bioinformatics 27:2987. https://doi.org/10.1093/BIOINFORMATICS/BTR509
976 Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler
977 transform. Bioinformatics 25:1754.
978 https://doi.org/10.1093/BIOINFORMATICS/BTP324
979 Li P, Quan X, Jia G, Xiao J, Cloutier S, You FM (2016) RGAugury: A pipeline for
980 genome-wide prediction of resistance gene analogs (RGAs) in plants. BMC
981 Genomics 17:1–10. https://doi.org/10.1186/S12864-016-3197-X
982 Liu J, Chen N, Grant JN, Cheng Z-M, Neal Stewart Jr C, Hewezi T (2015) Soybean
983 kinome: functional classification and gene expression patterns. J Exp Bot 66:1919–
984 1934. https://doi.org/10.1093/jxb/eru537
985 Liu Y, Li D, Yang N, Zhu X, Han K, Gu R, Bai J, Wang A, Zhang Y (2021) Genome-
986 wide identification and analysis of CC-NBS-LRR family in response to downy
987 mildew and black rot in Chinese cabbage. Int J Mol Sci 22:4266.
988 https://doi.org/10.3390/ijms22084266
989 Lozano R, Hamblin MT, Prochnik S, Jannink JL (2015) Identification and distribution
990 of the NBS-LRR gene family in the Cassava genome. BMC Genomics 16:1–14.
991 https://doi.org/10.1186/S12864-015-1554-9
992 Lozano R, Ponce O, Ramirez M, Mostajo N, Orjeda G (2012) Genome-wide
993 identification and mapping of nbs-encoding resistance genes in solanum tuberosum
994 group phureja. PLoS One 7. https://doi.org/10.1371/JOURNAL.PONE.0034775
995 Lu S, Wang J, Chitsaz F, …, Derbyshire M (2020) CDD/SPARCLE: the conserved
996 domain database in 2020. Nucleic Acids Res 48:D265–D268.

997 https://doi.org/10.1093/nar/gkz991
998 Lupas A, Van Dyke M, Stock J (1991) Predicting coiled coils from protein sequences.
999 Science (80-) 252:1162–1164. https://doi.org/10.1126/SCIENCE.252.5009.1162
1000 Ma Y, Chhapekar SS, Lu L, Oh S, Singh S, Kim CS, Kim S, Choi GJ, Lim YP, Choi SR
1001 (2021) Genome-wide identification and characterization of NBS-encoding genes in
1002 Raphanus sativus L. and their roles related to Fusarium oxysporum resistance.
1003 BMC Plant Biol 21:1–17. https://doi.org/10.1186/S12870-020-02803-8
1004 Macho AP, Zipfel C (2014) Plant PRRs and the activation of innate immune signaling.
1005 Mol Cell 54:263–272. https://doi.org/10.1016/J.MOLCEL.2014.03.028
1006 Meyers B, Kozik A, Griego A, … HK-TP, 2003 U (2003a) Genome-wide analysis of
1007 NBS-LRR–encoding genes in Arabidopsis. Plant Cell 15:809–834.
1008 https://doi.org/https://doi.org/10.1105/tpc.009308
1009 Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003b) Genome-wide
1010 analysis of NBS-LRR–encoding genes in Arabidopsis. Plant Cell 15:809–834.
1011 https://doi.org/10.1105/tpc.009308
1012 Milligan S, Bodeau J, Yaghoobi J, …, Kaloshian I (1998) The Root Knot Nematode
1013 Resistance Gene Mi from Tomato Is a Member of the Leucine Zipper, Nucleotide
1014 Binding, Leucine-Rich Repeat Family of Plant Genes. Plant Cell 10:1307–1319.
1015 https://doi.org/10.1105/tpc.10.8.1307
1016 Mitchell A, Attwood T, … BP (2019) InterPro in 2019: improving coverage,
1017 classification and access to protein sequence annotations. Nucleic Acids Res
1018 47:D351–D360. https://doi.org/10.1093/nar/gky1100
1019 Nascimento JIG, Stamford TCM, Melo NFCB, Nunes I dos S, Lima MAB, Pintado
1020 MME, Stamford-Arnaud TM, Stamford NP, Stamford TLM (2020) Chitosan–citric
1021 acid edible coating to control Colletotrichum gloeosporioides and maintain quality
1022 parameters of fresh-cut guava. Int J Biol Macromol 163:1127–1135.
1023 https://doi.org/10.1016/J.IJBIOMAC.2020.07.067
1024 Naseer S, Hussain S, Naeem N, Pervaiz M, Rahman M (2018) The phytochemistry and
1025 medicinal value of Psidium guajava (guava). Clin Phytoscience 4:1–8.
1026 https://doi.org/10.1186/S40816-018-0093-8
1027 Neupane S, Andersen E, Neupane A, Genes, Nepal M (2018) Genome-Wide
1028 Identification of NBS-Encoding Resistance Genes in Sunflower (Helianthus
1029 annuus L.). Genes (Basel) 9:384. https://doi.org/10.3390/genes9080384
1030 Nguyen L, Evolution, Schmidt H (2015) IQ-TREE: a fast and effective stochastic

1031 algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–
1032 274. https://doi.org/10.1093/molbev/msu300
1033 Paal JRÈ, Henselewski H, Muth J, Meksem K, Mene Â Ndez CM, Salamini F, Ballvora
1034 A, Gebhardt C (2004) Molecular cloning of the potato Gro1-4 gene conferring
1035 resistance to pathotype Ro1 of the root cyst nematode Globodera rostochiensis,
1036 based on a candidate gene. Plant J 38:285–297. https://doi.org/10.1111/j.1365-
1037 313X.2004.02047.x
1038 Pedersen T (2025) patchwork: The Composer of Plots. R package version 1.3.2.9000.
1039 https://patchwork.data-imaginist.com
1040 Porter BW, Paidi M, Ming R, Alam M, Nishijima WT, Zhu YJ (2009) Genome-wide
1041 analysis of Carica papaya reveals a small NBS resistance gene family. Mol Genet
1042 Genomics 281:609–626. https://doi.org/10.1007/S00438-009-0434-X
1043 Purwar S, Singh CM, Kumar M, Singh AK, Pratap A, Singh P, Gore PG, Singh NP
1044 (2023) Genome-Wide Identification and Analysis of NBS-LRR-Encoding Genes in
1045 Mungbean (Vigna radiata L. Wilczek) and Their Expression in Two Wild Non-
1046 progenitors Reveal Their Role in MYMIV Resistance. J Plant Growth Regul 1–14.
1047 https://doi.org/10.1007/S00344-023-10948-7
1048 Qiu Y (2024) showtext: Using Fonts More Easily in R Graphs.
1049 Quinton-Tulloch M, Agronomy, Steele K (2021) SNPs, indels, and microsatellites
1050 within and near to rice nbs-lrr resistance gene candidates. Agronomy.
1051 https://doi.org/10.3390/agronomy11112297
1052 Rahaman Khan M, Poornima K, Singh Somvanshi V, Kumar Walia R, Singh
1053 Somvanshi V, Kumar walia R (2022) Meloidogyne enterolobii: a threat to guava
1054 (Psidium guajava) cultivation. Arch Phytopathol Plant Prot 55:1961–1997.
1055 https://doi.org/10.1080/03235408.2022.2132623
1056 Ramírez-Zavaleta CY, García-Barrera LJ, Rodríguez-Verástegui LL, Arrieta-Flores D,
1057 Gregorio-Jorge J (2022) An Overview of PRR- and NLR-Mediated Immunities:
1058 Conserved Signaling Components across the Plant Kingdom That Communicate
1059 Both Pathways. Int J Mol Sci 23. https://doi.org/10.3390/IJMS232112974
1060 Ribeiro IJA, Pommer CV (2002) Breeding guava (Psidium guajava) for resistance to
1061 rust caused by Puccinia psidii. XXVI Int Hortic Congr Citrus Other Subtrop Trop
1062 Fruit Crop 75–78. https://doi.org/10.17660/ActaHortic.2004.632.9
1063 Roux F, Voisin D, Badet T, Balagué C, Barlet X, Huard-Chauveau C, Roby D, Raffaele
1064 S (2014) Resistance to phytopathogens e tutti quanti: placing plant quantitative

1065 disease resistance on the map. Mol Plant Pathol 15:427–432.
1066 https://doi.org/10.1111/mpp.12138
1067 Sagi MS, Deokar AA, Tar'An B (2017) Genetic analysis of NBS-LRR gene family in
1068 chickpea and their expression profiles in response to ascochyta blight infection.
1069 Front Plant Sci 8:838. https://doi.org/10.3389/FPLS.2017.00838/FULL
1070 Saha D, Rana R, Sureja A, Verma M, …, Arya L (2013) Cloning and characterization
1071 of NBS-LRR encoding resistance gene candidates from Tomato Leaf Curl New
1072 Delhi Virus resistant genotype of Luffa cylindrica. Physiol Mol Plant Pathol
1073 81:107–117. https://doi.org/10.1016/j.pmpp.2012.11.007
1074 Sansaloni C, Franco J, Santos B, Percival-Alwyn L, Singh S, Petroli C, Campos J,
1075 Dreher K, Payne T, Marshall D, Kilian B, Milne I, Raubach S, Shaw P, Stephen G,
1076 Carling J, Pierre C Saint, Burgueño J, Crosa J, Li HH, Guzman C, Kehel Z, Amri
1077 A, Kilian A, Wenzl P, Uauy C, Banziger M, Caccamo M, Pixley K (2020)
1078 Diversity analysis of 80,000 wheat accessions reveals consequences and
1079 opportunities of selection footprints. Nat Commun 11:1–12.
1080 https://doi.org/10.1038/S41467-020-18404-W;TECHMETA
1081 Sarris P, Duxbury Z, Huh S, Ma Y, C Segonzac C (2015) A plant immune receptor
1082 detects pathogen effectors that target WRKY transcription factors. Cell 161:1089–
1083 1100. https://doi.org/10.1016/j.cell.2015.04.024
1084 Saucet SB, Ghelder C Van, Abad P, Duval H, Esmenjaud D (2016) Resistance to root-
1085 knot nematodes Meloidogyne spp. in woody plants. New Phytol 211:41–56.
1086 https://doi.org/10.1111/nph.13933
1087 Savadi S, Sowmya K, Megha VS, Muralidhara BM, Mohana GS (2021) Genetic
1088 diversity and identification of interspecific hybrids of Anacardium species using
1089 microsatellites. Springer 44:139–148. https://doi.org/10.1007/S40415-020-00678-5
1090 Sela H, Spiridon LN, Petrescu AJ, Akerman M, Mandel-Gutfreund Y, Nevo E, Loutre
1091 C, Keller B, Schulman AH, Fahima T (2012) Ancient diversity of splicing motifs
1092 and protein surfaces in the wild emmer wheat (Triticum dicoccoides) LR10 coiled
1093 coil (CC) and leucine-rich repeat (LRR) domains. Mol Plant Pathol 13:276–287.
1094 https://doi.org/10.1111/J.1364-3703.2011.00744.X
1095 Sharma R, Rawat V, Suresh C (2017) Genome-wide identification and tissue-specific
1096 expression analysis of nucleotide binding site-leucine rich repeat gene family in
1097 Cicer arietinum (kabuli chickpea). Genomics data 14:24–31.
1098 https://doi.org/10.1016/j.gdata.2017.08.004

1099 Shi J, Zhang M, Zhai W, Meng J, …, Gao H (2018) Genome-wide analysis of
1100 nucleotide binding site-leucine-rich repeats (NBS-LRR) disease resistance genes in
1101 Gossypium hirsutum. Physiol Mol Plant Pathol 104:1–8.
1102 https://doi.org/https://doi.org/10.1016/j.pmpp.2018.07.007
1103 Sikandar A, Zhang MY, Wang YY, Zhu XF, Liu XY, Fan HY, Xuan YH, Chen LJ,
1104 Duan YX (2020) Nematodes a risk to agriculture. Appl Ecol Environ Res 18:1679–
1105 1690. https://doi.org/10.15666/aeer/1801_16791690
1106 Sinapidou E, Williams K, Nott L, Bahkt S, To È M, Crute ô I, Bittner-Eddy P, Beynon
1107 Ã J (2004) Two TIR:NB:LRR genes are required to specify resistance to
1108 Peronospora parasitica isolate Cala2 in Arabidopsis. Plant J 38:898–909.
1109 https://doi.org/10.1111/j.1365-313X.2004.02099.x
1110 Soewarto J, Hamelin C, Bocs S, Mournet P, Vignes H, Berger A, Armero A, Martin G,
1111 Dereeper A, Sarah G, Carriconde F, Maggia L (2019) Transcriptome data from
1112 three endemic Myrtaceae species from New Caledonia displaying contrasting
1113 responses to myrtle rust (Austropuccinia psidii). Data Br 22:794–811.
1114 https://doi.org/10.1016/j.dib.2018.12.080
1115 Stanke M, Steinkamp R, Waack S, Morgenstern B (2004) AUGUSTUS: A web server
1116 for gene finding in eukaryotes. Nucleic Acids Res 32:309–312.
1117 https://doi.org/10.1093/nar/gkh379
1118 Tarr DEK, Alexander HM (2009) TIR-NBS-LRR genes are rare in monocots: Evidence
1119 from diverse monocot orders. BMC Res Notes 2:1–10.
1120 https://doi.org/10.1186/1756-0500-2-197
1121 Tatusov RL, Koonin E V., Lipman DJ (1997) A genomic perspective on protein
1122 families. Science (80-) 278:631–637.
1123 https://doi.org/10.1126/science.278.5338.631
1124 Team P (2024a) RStudio: Integrated Development Environment for R. Boston,
1125 MAPosit Software. http://www.posit.co/
1126 Team RC (2024b) R: A Language and Environment for Statistical Computing.
1127 https://www.r-project.org/
1128 Timothy L. Bailey MB, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Noble WWL
1129 and WS (2009) MEME Suite: tools for motif discovery and searching. Nucleic
1130 Acids Res 37:W202–W208. https://doi.org/https://doi.org/10.1093/nar/gkp335
1131 TRIDGE (2025) Production of Fresh Guava.
1132 https://www.tridge.com/intelligences/guava/production. Accessed 19 Jun 2025

1133 Tsuda K, Katagiri F (2010) Comparing signaling mechanisms engaged in pattern-
1134 triggered and effector-triggered immunity. Elsevier 13:459–465.
1135 https://doi.org/10.1016/j.pbi.2010.04.006
1136 Van Der Vossen EAG, Rouppe Van Der Voort JNAM, Kanyuka K, Bendahmane A,
1137 Sandbrink H, Baulcombe DC, Bakker J, Stiekema WJ, Klein-Lankhorst RM
1138 (2000) Homologues of a single resistance-gene cluster in potato confer resistance
1139 to distinct pathogens: a virus and a nematode. Plant J 23:567–576.
1140 https://doi.org/10.1046/j.1365-313x.2000.00814.x
1141 Varshney RK, Mahendar T, Aggarwal RK, Börner A (2007) Genic molecular markers
1142 in plants: development and applications. Springer 1:13–29.
1143 https://doi.org/10.1007/978-1-4020-6295-7_2
1144 Vasimuddin M, Sanchit M, Heng L, Srinivas A (2019) Efficient architecture-aware
1145 acceleration of BWA-MEM for multicore systems. Proc - 2019 IEEE 33rd Int
1146 Parallel Distrib Process Symp IPDPS 2019 314–324.
1147 https://doi.org/10.1109/IPDPS.2019.00041
1148 Vos P, Simons G, Jesse T, Wijbrandi J, Heinen L, Hogers R, Frijters A, Groenendijk J,
1149 Diergaarde P, Reijans M, Fierens-Onstenk J, Both M de, Peleman J, Liharska T,
1150 Hontelez J, Zabeau M (1998) The tomato Mi-1 gene confers resistance to both
1151 root-knot nematodes and potato aphids. Nat Biotechnol 16:1365–1369.
1152 https://doi.org/https://doi.org/10.1038/4350
1153 Wang T, Jia Z, Zhang J, Liu M, Guo Z, G Wang G (2020) Identification and Analysis
1154 of NBS-LRR Genes in Actinidia chinensis Genome. Plants 9:1350.
1155 https://doi.org/10.3390/plants9101350
1156 Wang Y, Song Z, Zhang W, Xu L, Su X, …, Muleke E (2017) Identification and
1157 characterization of expressed TIR-and non-TIR-NBS-LRR resistance gene
1158 analogous sequences from radish (Raphanus sativus L.) de novo. Sci Hortic
1159 (Amsterdam) 216:284–292. https://doi.org/10.1016/j.scienta.2016.12.019
1160 Wei H, Li W, Sun X, Zhu S, Zhu J (2013) Systematic Analysis and Comparison of
1161 Nucleotide-Binding Site Disease Resistance Genes in a Diploid Cotton Gossypium
1162 raimondii. PLoS One 8:e68435.
1163 https://doi.org/10.1371/JOURNAL.PONE.0068435
1164 Wickham H, Averick M, Bryan J, Chang W, D' L, Mcgowan A, François R, Grolemund
1165 G, Hayes A, Henry L, Hester J, Kuhn M, Lin Pedersen T, Miller E, Bache SM,
1166 Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D,

1167 Wilke C, Woo K, Yutani H (2019) Welcome to the Tidyverse. joss.theoj.org
1168 4:1686. https://doi.org/10.21105/JOSS.01686
1169 Win J, Chaparro-Garcia A, Belhaj K, Saunders DGO, Yoshida K, Dong S, Schornack S,
1170 Zipfel C, Robatzek S, Hogenhout SA, Kamoun S (2012) Effector biology of plant-
1171 associated organisms: Concepts and perspectives. Cold Spring Harb Symp Quant
1172 Biol 77:235–247. https://doi.org/10.1101/SQB.2012.77.015933
1173 Xiao F, Hu J, Xu Z, Liao K, Agricultural TL-J of C, 2010 U (2010) Cloning and
1174 expression analysis of fragments related to root-knot nematode resistance from
1175 wild Myrobalan plum (Prunus sogdiana). J China Agric Univ 15:77–83
1176 Zhang H, Zhang J, Xu Q, Wang D, Di H, Huang J, Yang X, Wang Z, Zhang L, Dong L,
1177 Wang Z, Zhou Y (2020a) Identification of candidate tolerance genes to low-
1178 temperature during maize germination by GWAS and RNA-seqapproaches. BMC
1179 Plant Biol 20:1–17. https://doi.org/10.1186/s12870-020-02543-9
1180 Zhang W, Yuan Q, Wu Y, Zhang J, Of JN-IJ, 2022 U (2022) Genome-Wide
1181 Identification and Characterization of the CC-NBS-LRR Gene Family in
1182 Cucumber (Cucumis sativus L.). Int J Mol Sci 23:5048.
1183 https://doi.org/10.3390/ijms23095048
1184 Zhang YM, Chen M, Sun L, Wang Y, Yin J, Liu J, Sun XQ, Hang YY (2020b)
1185 Genome-Wide Identification and Evolutionary Analysis of NBS-LRR Genes From
1186 Dioscorea rotundata. Front Genet 11:484.
1187 https://doi.org/10.3389/FGENE.2020.00484/FULL
1188 Zhao T, Miao L, Zou M, Hussain I, Yu H, Li J, Sun N, Kong L, Wang S, Li J, Yu X
1189 (2024) Development of SSRs Based on the Whole Genome and Screening of
1190 Bolting-Resistant SSR Marker in Brassica oleracea L. Horticulturae.
1191 https://doi.org/10.3390/horticulturae10050443
1192 Zhong Y, Chen Z, Cheng ZM (2022) Different scales of gene duplications occurring at
1193 different times have jointly shaped the NBS-LRR genes in Prunus species. Mol
1194 Genet Genomics 297:263–276. https://doi.org/10.1007/S00438-021-01849-Z
1195 Zhou T, Wang AY, Chen J-Q, Araki AH, Jing AZ, Jiang K, Shen AJ, Tian AD (2004)
1196 Genome-wide identification of NBS genes in japonica rice reveals significant
1197 expansion of divergent non-TIR NBS-LRR genes. Mol Genet Genomics 271:402–
1198 415. https://doi.org/10.1007/s00438-004-0990-z
1199

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial1.xlsx
- Supplementarymaterial2.xlsx
- Supplementarymaterial3.tif
- Supplementarymaterial4.fna
- Supplementarymaterial5.txt
- Supplementarymaterial6.txt
- Supplementarymaterial7.csv