

Genome characterization and identification of leucine-rich repeat receptor (LRR) genes for pathogen resistance in wild banana *Musa acuminata* subsp. *halabanensis*

Tamimah Shafwatul Ishlah¹, I Made Artika¹*, Rahadian Pratama¹, and Fajarudin Ahmad²

¹Department of Biochemistry, IPB University, Indonesia

²Research Center for Applied Botany, National Research and Innovation Agency, Indonesia

Abstract. Bananas are one of the most widely consumed fruits in the world. However, pathogenic diseases cause significant losses in cultivation. Leucine-rich repeat (LRR), encoded by resistance genes, is a protein domain that plays a role in the innate immune system against pathogens. Identifying resistance genes in wild bananas is expected to aid in cultivating disease-resistant banana cultivars. Resistance gene studies have not yet been conducted on the wild banana *Musa acuminata* subsp. *halabanensis* (Meijer) M.Hotta. This study aimed to characterize the genome and identify LRR resistance genes and their defense mechanisms in *M. acuminata* subsp. *halabanensis*. Quality control was performed on raw sequencing data from Oxford Nanopore Technologies (ONT) and Illumina, followed by sequence hybrid assembly and annotation. The results showed that the hybrid assembly of ONT sequences and polishing with Illumina improved sequence quality, with a Q-score >30 reaching 90,12%. The assembly produced 9.793 contigs, with a total banana genome size of 437.161.969 bp. Functional annotation identified 16 genes belonging to the LRR group, including *lrr1*, *bak1_1,4,5*, *bri1_1-4*, *ghr1_1,2,4*, *pir1_1,3,4,6* and *serk4*. The LRR mechanism in pathogen defense was demonstrated by the activation of defense signalling pathways

1 Introduction

Indonesia harbours a rich diversity of banana species, including cultivated and wild species. One wild banana species identified in Sumatra, Indonesia, is *Musa acuminata* subsp. *halabanensis* [1]. This wild banana is known to be a primary progenitor of cultivated bananas such as the Madu banana (*Musa* spp.) [2]. However, banana cultivation faces significant losses annually due to pathogenic diseases.

¹ Corresponding author: imart@apps.ipb.ac.id

Various banana accessions have been identified to possess resistance genes (R) against pathogens. Pathogenesis-related (PR) genes have been found in cultivated bananas *M. acuminata* (AAA group) 'Khai Kasetsart 2' and 'Raksa,' conferring resistance to bacterial wilt [3]. In wild *M. acuminata* subsp. *burmannicoides* var. *Calcutta* and cultivated *Musa* sp. 'Prata-anã,' the TL1 (Traumat-in-like protein) gene, has been detected in the roots as a defence mechanism against *Fusarium* wilt [4]. Furthermore, wild *M. acuminata* subsp. *malaccensis* and *M. balbisiana* are known to possess the *MaNBS89* gene for protection against *Fusarium* [5].

Most R genes are known to encode Leucine-Rich Repeat (LRR) domains, which are found in the receptor protein kinase (RPK) and nucleotide-binding site-leucine-rich repeat (NBS-LRR) protein families [6]. A previous study identified LRR-RLP encoding genes belonging to the RPK protein group in banana *Musa acuminata* subsp. *malaccensis*. Meanwhile, in *Musa acuminata* Colla var. *malaccensis* (Ridl.) Nasution and *Musa acuminata* AAA group 'Cavendish,' genes encoding LRRs classified into the non-TIR-NBS-LRR group have been identified [7]. The LRR domain plays a role in the innate immune system by recognizing specific molecular patterns of pathogens and activating immune signalling pathways to combat them. Therefore, LRR-encoding genes can be used as target genes for R gene identification.

Target gene identification can be achieved through various approaches. Whole Genome Sequencing (WGS) facilitates comprehensive genome mapping and enables the identification of resistance genes in bananas by providing access to the *Musa* reference genome, as well as allowing the detection of single nucleotide polymorphisms (SNPs), analysis of genome evolution, and insights into population genetics [8]. However, despite its role as a genetic contributor to cultivated bananas, resistance gene identification has not yet been conducted on the wild banana *M. acuminata* subsp. *halabanensis*. Therefore, this study aims to characterize its genomic information using WGS and to identify genes associated with resistance to bacterial wilt and *Fusarium* wilt pathogens in *M. acuminata* subsp. *halabanensis*.

2 Material and Methods

2.1 Material

This study utilized raw DNA sequencing data generated using ONT-PromethION and Illumina-Novaseq 6000 from a wild banana accession, *Musa acuminata* subsp. *halabanensis* (Meijer) M.Hotta, originally from Riau and cultivated at the Cibinong Botanical Garden Collection, BRIN (collection no. 257b; accession no. B2021110012). The plant tissue used for DNA extraction was the folded leaf shoot (young rolled leaf). This study also incorporated homologous gene sequence data from members of the Musaceae family.

2.2 Methods

2.2.1 QC and Filtering of Sequencing Data

Quality control (QC) and filtering were conducted on sequence data obtained from ONT and Illumina platforms. All analyses were performed on the UseGalaxy platform (<https://usegalaxy.org>). The ONT sequence data were analyzed using the NanoPlot v1.42.0, while the Illumina sequence data were analyzed using Fastp v0.20.1. The option `-read_type 1D` was selected to specify the read type of the raw reads generated from ONT. Plot

construction was selected with `--color` and `--plots` customized as needed. Additionally, the `--N50` option was activated. The options selected for raw reads generated from Illumina were : `-q 20 -l 100 --detect_adapter_for_pe -f 5 -t 1 -F 5 -T 1`. The NanoPlot and Fastp results were assessed for reads performance.

2.2.2 Hybrid Assembly of ONT and Illumina Sequencing Data

The ONT sequence data were assembled using Flye v2.9.3. The selected parameter options were : `--nano-raw`, `--iterations-int 1`, `--keep-haplotypes yes`, `--scaffold yes`, and `--meta no`. Then, the draft assemblies from ONT data were polished using MEDAKA v1.7.2 with basecalling set to "raw data", and the model `r941_min_high_g360` was selected.

The polished assemblies were further corrected using QC-filtered Illumina short reads with Pilon v1.20. The ONT contig output and Illumina reads were used as input and this input type was automatically set to BAM, the option `--vcf no` was selected, and all other parameters were left at their default settings.

Corrected contig statistics were evaluated using QUAST v5.2.0. The input was set to individual assembly (1 contig file/sample) using the corrected ONT assembly output. Parameters were configured as : `--assembly-type genome`, `--eukaryote yes` and `-output; HTML report, Tabular report, Log file, PDF file`. All tools in this analysis were performed on the UseGalaxy platform (<https://usegalaxy.org>).

2.2.3 Gene Structure Annotation and Characterization

Structural prediction and gene targeting were conducted using BRAKER3 v3.0.8. BRAKER was executed with relevant command line options such as : `--genome=genome.fa`, `--species=speciesname`, `--crf yes`, and the `--AUGUSTUS_ab_initio` output was saved in the file `augustus.ab_initio.gtf`. The resulting gene prediction output was then analyzed to determine gene function using Funannotate Functional v1.8.15. The selected parameter options were : `-m diamond`, `-i braker/braker.aa`, `--dmnd_db /ext-db/eggno/viridiplantae.dmnd`, `-o halabanensis`, `--output_dir eggno`, `--dbmem`, and `--cpu 10`. The tabular output file was used for functional analysis of LRR target genes. BRAKER3 and Funannotate Functional were performed on the UseGalaxy platform (<https://usegalaxy.org>).

2.2.4 LRR Gene Analysis

Predicted genes or proteins were presented in the tabular output of the eggNOG-mapper. Clusters of Orthologous Groups (COG) were summarized and visualized in charts using Excel. The R gene groups that encode LRR domains were compiled. LRR domain structural analysis was carried out using InterPro. In parallel, the nucleotide sequences of LRR genes were BLASTed against reference genomes using nucleotide queries (BLASTn) within the Musaceae family. Homologous sequences for each accession were downloaded in FASTA format for phylogenetic tree construction.

2.2.5 Phylogenetic Analysis

Phylogenetic analysis was conducted using MEGA X v10.2.2 on predicted LRR-encoding genes and their homologous sequences. Multiple sequence alignment was performed using ClustalW with default parameters. Phylogenetic trees were constructed using the neighbor-joining algorithm with 1000 bootstrap replicates to assess tree topology reliability. Branches with bootstrap values >70% were retained. The final phylogenetic tree was exported in

Newick format (.nwk) and uploaded to the iTOL web server to generate a cladogram-style phylogenetic tree.

2.2.6 Defense Mechanism Analysis

Defense mechanisms of protein LRR-domains encoded by the target genes were investigated using the KEGG pathway. KO database results from the annotation were searched in KEGG to analyze relevant signalling pathways. The defense mechanism was visualized using BioRender (<https://www.biorender.com/>).

3 Results

3.1 Characteristics of Sequencing Data and Genome of *Musa acuminata* subsp. *halabanensis*

The quality and statistics of the sequencing reads were obtained through quality control (QC) and filtering processes. A comparison of ONT and Illumina sequencing QC results is presented in Table 1. These results illustrate the differences in data characteristics generated by long-read (ONT) and short-read (Illumina) sequencing technologies. Based on the mean read length and Q-score values, ONT produced longer reads with fewer total sequences compared to Illumina. However, Illumina reads exhibited higher quality than ONT.

Table 1. Comparison of ONT and Illumina sequencing data characteristics based on QC analysis.

Parameters	ONT	Illumina	
		Before Filtering	After Filtering
Total Reads	5.750.000	127.931.696	123.600.492
Total Base (bp)	15.553.832.704	19.061.823.000	16.883.571.000
Mean Read Length (bp)	2.705	149	136
Q-score (%)	>Q15 = 53,2	>Q20 = 92,89 >Q30 = 87,72	>Q20 = 94,86 >Q30 = 90,12
N50 (bp)	5.005	-	-
GC%	-	39,48	38,70

The QC and filtering process is a crucial step in any data analysis, as it helps identify potential biases and technical errors during sequencing, which could affect downstream analyses and result interpretation. QC analysis of the ONT sequencing data revealed that read lengths for *M. acuminata* subsp. *halabanensis* ranged from 0 to 60.000 bp (Figure 1a) with the longest read reaching 161.145 bp. The average quality scores ranged from Q5 to Q15 and highest quality reads >Q35, with relatively few numbers. The minimum quality parameters for the sequencing data included a read length of ≥500 bp and a read quality score (Q-score) >Q7 [9]. In total, 5.749.983 reads passed the quality threshold (>Q7) and the mean read lenght of 2.705 bp. These results indicated that the ONT sequencing data have met the quality criteria. Although a few reads reached the highest quality score of Q35, most of the reads passed the quality threshold of Q7. Therefore, this suggested that the sequencing data generated using ONT has good read quality and length, making it suitable for further analysis.

Illumina sequencing was used to enhance the completeness and accuracy of the genome assembly. This analysis was based on the paired-end sequencing strategy, which relies on overlapping regions between read pairs. Non-overlapping read pairs were likely due to fragments that were either too short (<30 bp), too long (>268 bp), or contained too many sequencing errors to be correctly aligned. The average read length detected during sequencing in this study was 249 bp, as shown in the insert size, which fell within the appropriate fragment size range. It indicated that the detected reads generally overlap, allowing for the generation of reliable sequence assemblies.

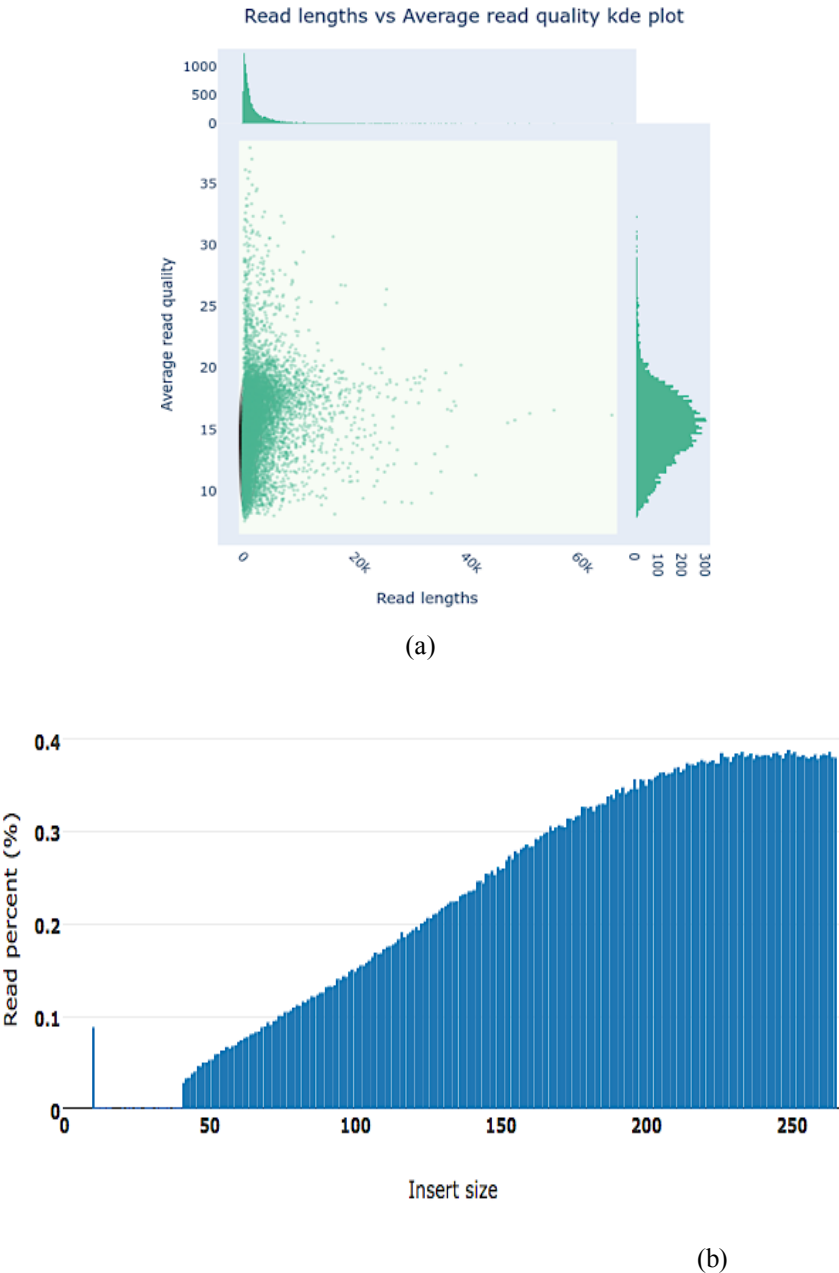
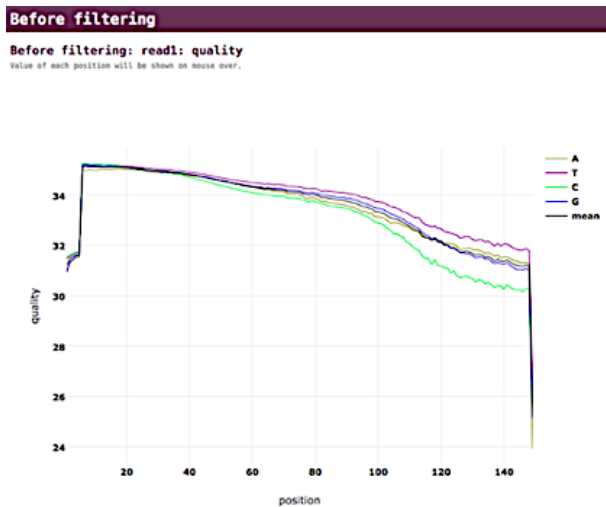


Fig. 1. Histogram of read fragment distribution based on sequencing QC results. (a) Read length and quality distribution of ONT reads generated using NanoPlot, ranging from 0 to 60,000 bp with an

average quality between Q5–Q15; (b) The average insert size of Illumina reads detected by Fastp was 249 bp.

The QC procedures included removing adapter contamination, filtering low-quality reads, excluding undetermined bases ('N'), and correcting base-calling errors. This process revealed differences in read quality before and after filtering (Figure 2). The graph displays five colors representing the four nucleotide bases and their average quality after filtering. Before filtering, a full-quality peak is visible across all read positions, indicating raw base quality values. After filtering, global trimming was applied by removing five bases from the 5' end and one base from the 3' end, resulting in a smoothed sequence quality plot without distinct peaks. During sequencing, some terminal thymine (T) and cytosine (C) bases can be misread as guanine (G), resulting in a poly-G tail. This region was trimmed to improve data quality. Furthermore, approximately 0.001% of unidentified bases ('N') present before filtering were eliminated. High-quality reads are indicated by base quality scores above Q30. In this study, Illumina reads showed good sequencing quality, with base quality scores exceeding Q30, reaching an average of 34. These results showed that the ONT and Illumina sequencing data passed quality control.



(a)

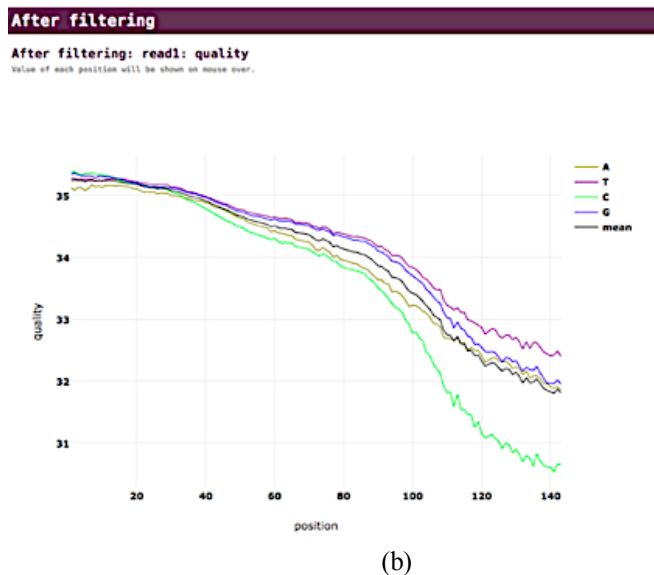


Fig. 2. Quality filtering graph of Illumina sequencing data. (a) Before filtering, all quality peaks are visible; (b) After filtering, low-quality reads are removed, resulting in read quality scores above Q30.

Based on the results obtained, ONT reads showed longer read lengths but fewer total reads than Illumina. However, Illumina reads had higher quality than ONT reads. Therefore, both sequencing technologies were combined to build the complete genome of *Musa acuminata* subsp. *halabanensis*. High-quality ONT and Illumina sequencing reads were assembled into contiguous sequences known as contigs. The ONT long-read contigs were polished using Illumina short reads to correct base errors, thereby improving the genome assembly's accuracy, continuity, and completeness. The characteristics of the assembled genome are shown in table 2, with an estimated genome size of 437.16 Mb and a total of 9,793 contigs. Compared to other wild *M. acuminata* subspecies such as *M. acuminata* subsp. *malaccensis* (450.43 Mb), *zebrina* (550.9 Mb), *banksii* (484.8 Mb), and *burmannica* (505.1 Mb), the genome of *M. acuminata* subsp. *halabanensis* was shorter [2,10]. In contrast, *Musa acuminata* subsp. *halabanensis* have shown as a major contributor to one of the haplotypes (H1) of the cultivated banana ‘Madu’ (which has haplotypes H1 and H2). The sequencing and assembly processes carried out separately for each haplotype of the ‘Madu’ banana revealed that the genome size of H1 reached 529.6 Mb [2]. It demonstrates that using either cultivated or wild banana samples for sequencing can result in significant differences in genome size. The choice of assembly and polishing tools also influences such variation. Nevertheless, this study on genome size and characteristics provide valuable initial information for the genomic characterization of the wild banana *M. acuminata* subsp. *halabanensis*.

Parameters	Hybrid Assembly: ONT + Illumina (Quast)
Contig	9.793
Genom Size (bp)	437.161.969
N50	426.589
GC Content (%)	38.99

Table 2. Genome characteristics of *M. acuminata* subsp. *halabanensis* after polishing

3.2 Characteristics of Resistance Genes Encoding LRR Protein

The annotation process carried out the identification and characterization of R genes encoding protein LRR-domains. One functional annotation result was the Clusters of Orthologous Groups (COG) data, which classifies proteins based on their biological functions and homology into specific categories. The corresponding encoded proteins were classified as visualized in a COG diagram. The focus of this study was on LRR protein groups. Based on their biological functions, LRR proteins were categorized into groups involved in lipid transport and inorganic ion metabolism (categories I and P), secondary metabolite catabolism (Q), signal transduction mechanisms (T), intracellular trafficking and vesicular transport (U), and proteins with unknown functions (S) (Figure 3). The category T was identified as the primary biological role [34]. Nevertheless, some of these proteins are still classified under categories with unknown functions (S). This is likely due to the limitations of the current database information that have not yet been extensively explored. These proteins, despite being poorly understood, could be crucial for plant immunity and warrant further research, especially in the context of plant defense mechanisms.

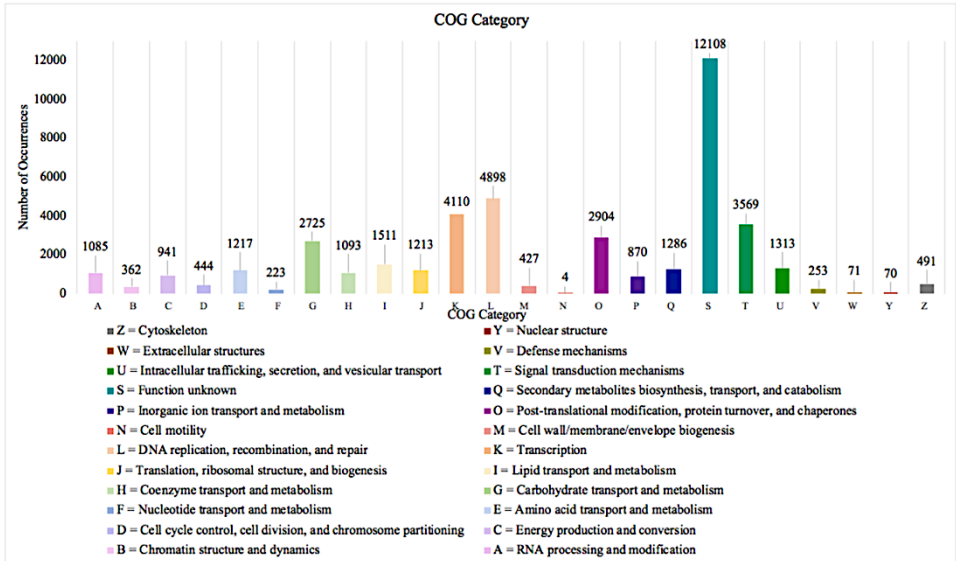


























Fig. 3. COG categories of annotated proteins based on biological functions and evolutionary homology. LRR proteins are classified I, P, Q, S, T, and U.

A total of 16 LRR-encoding genes were identified, including *lrr1*, BRI1-associated receptor kinase (*bak1_1,4,5*), Brassinosteroid Insensitive 1 (*bri1_1-4*), Guard Cell Hydrogen Peroxide-Resistant1 (*ghr1_1,2,4*), Plant Intracellular Ras-group related repeat proteins (*pir1,3,4,6*), and Somatic Embryogenesis Receptor Kinase (*serk4*). Protein domain analysis based on their amino acid sequences indicated that each protein possessed unique repetitive LRR domains (Table 3). All 16 genes were classified as resistance (R) genes encoding LRR domains within the receptor-like kinase (RPK) protein family. Specifically, *BAK1_1*, *BAK1_4*, *BAK1_5*, *BRI1_2*, and *BRI1_4* belong to the LRR receptor-like serine/threonine protein kinase (LRR-RLK) group; *BRI1_1*, *BRI1_3*, *GHR1_1*, *GHR1_2*, and *GHR1_4* fall under serine/threonine protein kinase (STK); *LRR1* and *SERK4* are part of the somatic embryogenesis receptor kinase (SERK) group; and *PIRL1*, *PIRL2*, *PIRL3*, and *PIRL4* are classified as receptor-like proteins (RLP).

Table 3. R genes encoding LRR proteins and their domain diversity

Protein Group	R Gene	Domains	Protein Group	R Gene	Domains
LRR-RLK	<i>bak1_1</i>		STK	<i>bri1_1</i>	
	<i>bak1_4</i>			<i>bri1_3</i>	
	<i>bak1_5</i>			<i>ghr1_1</i>	
	<i>bri1_2</i>			<i>ghr1_2</i>	
	<i>bri1_4</i>			<i>ghr1_4</i>	
LRR-RLP	<i>pir11</i>		LRR-RLP	<i>pir14</i>	
	<i>pir13</i>			<i>pir16</i>	
SERK	<i>lrr1</i>		SERK	<i>serk4</i>	

Symbol :  = LRR  = LRR type-8  = Transmembran
 = LRR-N terminus (LRRNT)  = LRR type-14  = BR11
 = LRR type-2  = Protein Kinase

The proteins encoded by the R genes contain characteristic and repetitive leucine-rich repeat (LRR) domains composed of distinct amino acid sequences. Proteins categorized as LRR-RLKs consist of both an LRR domain and an intracellular kinase domain. Typically, these proteins bind extracellular signaling molecules via the LRR domain, activating the intracellular kinase domain [11]. In contrast, proteins classified as LRR-RLPs possess only the LRR domain without a kinase domain. Nevertheless, they can still interact with other proteins to perform signaling functions. Proteins in the STK group generally contain both an LRR domain and a kinase domain with serine/threonine residues. The intracellular kinase domain can activate the mitogen-activated protein kinase (MAPK) signaling pathway in response to interactions between the LRR domain and pathogen molecules [6]. Meanwhile, SERK proteins have domains similar to those of LRR-RLKs and LRR-RLPs. In *Arabidopsis thaliana*, the SERK protein group functions as co-receptors in various signaling pathways, including those involved in morphogenesis, phytohormone signaling, and stress responses [11].

3.2 Phylogenetic Analysis

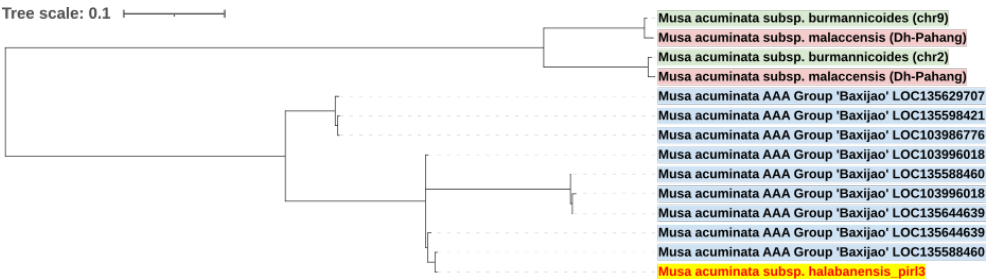
Phylogenetic analysis examined the evolutionary relationship of LRR-encoding genes in *M. acuminata* subsp. *halabanensis* with those from other banana species. The results showed that *lrr1* and *bri1_1* exhibited low bootstrap values (10% and 53%), indicating weak genetic relatedness to the cultivar *M. acuminata* AAA Group 'Baxijao.' In contrast, *pir13* displayed a high bootstrap value of 95%, signifying substantial genetic similarity to the same cultivar. Despite varying bootstrap support, these three genes clustered within the same clade as 'Baxijao,' suggesting a common ancestor or genetic relationship (Figure 4). The *serk4* (43%) and *bak1_1* (4%) also showed weak relatedness to wild banana *M. acuminata* subsp. *malaccensis*, while *ghr1_1* (87%) demonstrated substantial genetic similarity with *M. acuminata* var. *zebrina* on chromosome 4.

Based on its dispersal pattern, bananas originated from the northern Indo-Burma region (northeastern India, northern Myanmar, and China), then spread into two groups. Group 1 moved eastward toward Borneo and Papua (Indonesia), while Group 2 dispersed toward southern Indo-Burma (southern Myanmar, Thailand, Laos, and Vietnam), Sumatra–Malayan Peninsula, and southern China. The cultivated dessert banana cultivar 'Baxijao' is known to

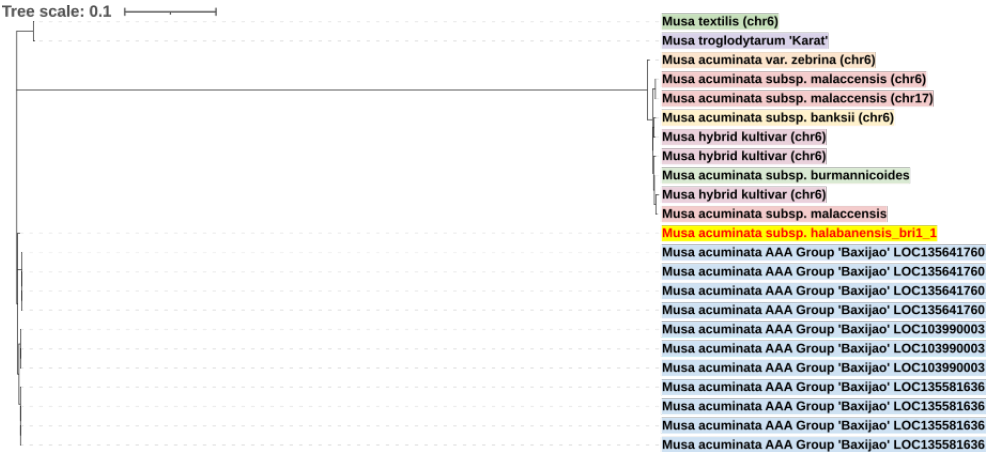
originate from the Hainan region in southern China. This cultivar belongs to the AAA genomic group derived from *Musa acuminata* and is classified within the *EuMusa* lineage. *EuMusa* is part of Group 2 in its dispersal pathway, which may explain the genetic similarities between ‘Baxijiao’ and *M. acuminata* subsp. *halabanensis*, found in Riau, Sumatra. Similar genetic resemblance is also observed with wild bananas, such as *M. acuminata* subsp. *malaccensis* and *M. acuminata* var. *zebrina*, which are also distributed within the Group 2 region [12].



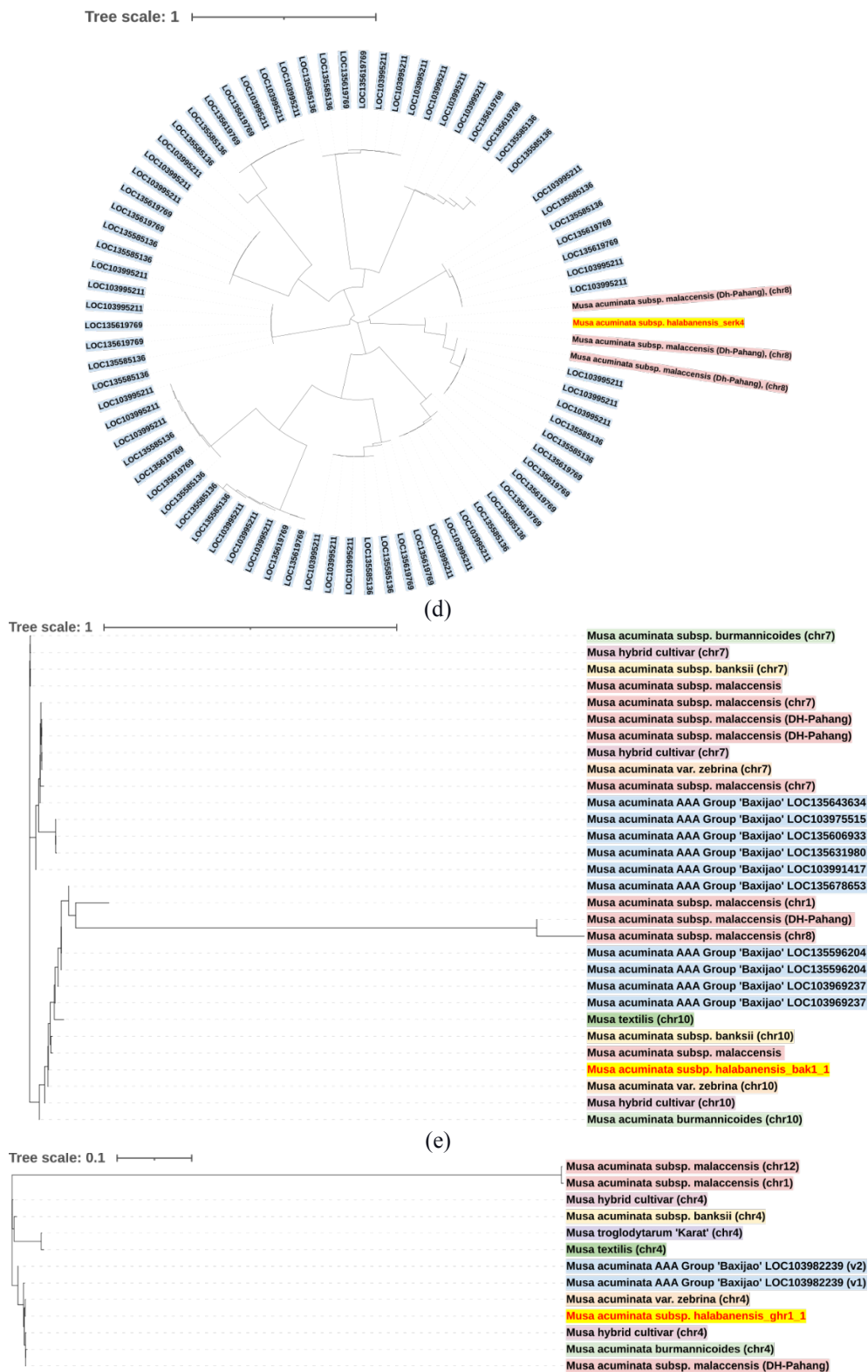
(a)



(b)



(c)



(f)

Fig. 4. Phylogenetic tree of LRR genes and their homologs across *Musa* species. Circular icons indicate abundant homologs, and blue-colored accessions represent the cultivar *M. acuminata* AAA Group 'Baxijao'. Genetic relationships are illustrated for: (a) *lrr1*, (b) *pir13*, and (c) *bri1_1* with 'Baxijao'; (d) *serk4* and (e) *bak1_1* with *M. acuminata* subsp. *malaccensis*; and (f) *ghr1_1* with *M. acuminata* var. *zebrina*.

3.4 Defense Mechanism of LRR Genes

The defense mechanism mediated by LRR genes was inferred from functional annotation results in the KO database. In this study, BAK1 (*K13416*, map 04626), SERK4 (*K13417*, map 04626) and members of the RLK protein family were identified and mapped in the KEGG pathway. These proteins, encoded by R genes, were involved in the plant-pathogen interaction pathway, specifically within the PAMP-triggered immunity (PTI) mechanism, which includes mitogen-activated protein kinase (MAPK) signaling cascades as co-receptors (Figure 5). PTI is an immune system activated by molecular patterns such as chitin from fungi, flagellin from bacteria, or damage-associated patterns including peptides and oligosaccharides generated during plant-pathogen interactions [13]. Pattern recognition receptors (PRRs), receptor proteins on the plasma membrane, recognize these patterns. The FLS2 (flagellin sensing 2) and EFR (EF-Tu receptor), known as PRRs, interact with co-receptors BAK1, SERK4, or other RLKs and activate MAPK signals [11].

MAPK is an immune signal transduction pathway that activates phosphorylation cascades, leading to changes in gene expression and cellular activity [14]. Based on the observed mechanism, the defense response is initiated through the activation of MEKK1, which subsequently phosphorylates two parallel MAPKK pathways, MKK1/2 and MKK4/5. MKK1/2 activates, while MKK4/5 activates MPK3/6. The activation of MPK3, MPK4 and MPK6 plays a crucial role in initiating the plant immune response by activating transcription factors. The MPK4 pathway leads to the activation of WRKY25, whereas the MPK3/6 pathway activates WRKY29. These transcription factors function in the nucleus to induce the expression of defense-related genes such as Flg22-induced Receptor-like Kinase 1 (*FRK1*), Pathogenesis-Related protein 1 (*PR1*), *WRKY29*, and Nonhost Resistance 1 (*NHO1*), which are involved in phytoalexin biosynthesis, salicylic acid (SA)-mediated immunity, miRNA accumulation, and nonhost pathogen defense responses. In *Arabidopsis*, the MEKK1-MKK1/2-MPK4 cascade suppresses immune responses, thereby preventing overactivation of the immune system. This pathway maintains homeostasis between plant defense and growth, ensuring a balanced immune response [15].

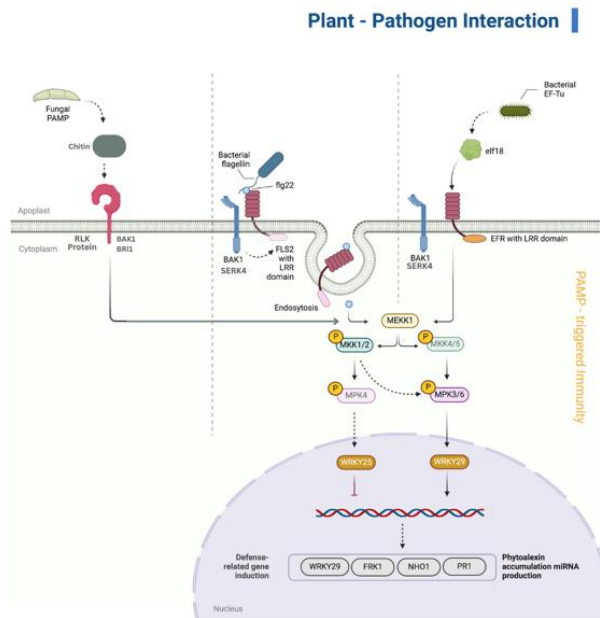


Fig. 5. The defense mechanism of BAK1_1, SERK4, and RLK proteins in recognizing pathogen effectors and activating PTI and MAPK signaling pathways, leading to the expression of defense-related genes. This figure was created using BioRender (<https://biorender.com/>).

4 Conclusion

The wild banana *Musa acuminata* subsp. *halabanensis* has a genome size of 437,161,969 bp with 9,793 contigs assembled from ONT and Illumina sequencing data. Functional annotation revealed 16 genes encoding LRR-domain proteins—including *lrr1*, *bak1_1,4,5*, *bri1_1-4*, *ghr1_1,2,4*, *pir1_1,3,4,6*, and *serk4*—that are genetically related to both wild and cultivated banana accessions. These LRR proteins are predominantly involved in signal transduction pathways such as PTI and MAPK, which are central to pathogen defense responses. The identification of these defense-related genes provides a valuable genetic resource for banana breeding programs, especially in developing disease-resistant cultivars. Moreover, understanding the genetic basis of pathogen recognition and signaling in wild banana offers a foundation for future research into molecular mechanisms of plant immunity. Further studies involving advanced sequencing technologies, functional analysis of uncharacterized genes (category S), and in vitro validation will enhance our capacity to exploit wild germplasm for sustainable banana improvement.

Acknowledgement

We express our gratitude to the Department of Biochemistry, IPB University, and the Life Sciences and Environmental Research Organization Program of the National Research and Innovation Agency (BRIN) for funding under the 2022 fiscal year project titled "Exploration of Genetic Diversity in Wild Bananas *M. acuminata* ssp. *halabanensis*, ssp. *sumatrana*, and ssp. *malaccensis* Using WGS."

References

1. T. Itino, M. Kato, M. Hotta, Pollination Ecology of the Two Wild Bananas, *Musa acuminata* subsp. *halabanensis* and *M. salaccensis*: Chiropterophily and Ornithophil. *Biotropica*. **23**, 151–158 (1991). <https://doi.org/10.2307/2388300>
2. G. Martin, B. Istace, F.C. Baurens, C. Belser, C. Hervouet, K. Labadie, C. Cruaud, B. Noel, C. Guiougou, F. Salmon, J. Mahadeo, F. Ahmad, H.A. Volkaert, G. Droc, M. Rouard, J. Sardos, P. Wincker, N. Yahiaoui, J.A. Aury, A. D'Hont, Unravelling genomic drivers of speciation in *Musa* through genome assemblies of wild banana ancestors. *Nat Commun*. **16**, 1-14 (2025). <https://doi.org/10.1038/s41467-025-56329-4>
3. J. Nitayaros, T. Thanyasiriwat, A. Sangdee, L. Rattanapolsan, R. Boonruangrod, P. Kawicha, Evaluation of banana cultivars and the pathogenesis-related class 3 and 10 proteins in defense against *Ralstonia syzygii* subsp. *celebesensis*, the causal agent of banana blood disease. *J Plant Prot Res*. **63**, 375–386 (2023). <https://doi.org/10.24425/jppr.2023.146873>
4. É. de C. Costa, L.S. Bastos, T.G. Gomes, R.N.G. Miller RNG, Reference genes for RT-qPCR analysis in *Musa acuminata* genotypes contrasting in resistance to *Fusarium oxysporum* f. sp. *cubense* subtropical race 4. *Sci Rep*. **14**, 1-11 (2024). <https://doi.org/10.1038/s41598-024-67538-0>
5. Y. le Huo, S. wen Liu, H. qing Huang, Z. yuan Li, M. Ahmad, M. xia Zhuo, C. yu Li, B. Liu, Y. dong Li, NBS-LRR genes of *Musa acuminata* is involved in disease resistance to *Fusarium* wilt. *Sci Hortic*. **336**, 1-10 (2024). <https://doi.org/10.1016/j.scienta.2024.113361>
6. P. Arya, V.Acharya. Plant STAND P-loop NTPases: a current perspective of genome distribution, evolution, and function: Plant STAND P-loop NTPases: genomic organization, evolution, and molecular mechanism models contribute broadly to plant pathogen defense. *Molecular Genetics and Genomics*. **293**, 17–31 (2018). <https://doi.org/10.1007/s00438-017-1368-3>
7. D. Martanti, U. Widyastuti, R. Y.S. Poerba, Megia, Identification of Gene Candidate of Nucleotide Binding Site (NBS) from Banana *Musa acuminata* Colla var. *malaccensis* (Riddl.) Nasution and *Musa*, AAA, Cavendish Sub-group. *Pak. J. Biol. Sci*. **18**, 99-106 (2015). <https://doi.org/10.3923/pjbs.2015.99.106>
8. J. Sardos, X. Perrier, J. Doležel, E. Hříbová, P. Christelová, I. V. den Houwe, A. Kilian, N. Roux, DARt whole genome profiling provides insights on the evolution and taxonomy of edible Banana (*Musa* spp.). *Ann Bot*. **118**, 1269–1278 (2016). <https://doi.org/10.1093/aob/mcw170>
9. A.A.E. Kristianti, F.G. Dwiyaniti, R. Pratama R, I.Z. Siregar, Characterization of the chloroplast genome sequence of *Calophyllum inophyllum*, a bioenergy tree species, using Oxford Nanopore Technologies. *IOP Conf. Ser.: Earth Environ. Sci*. **1315**, 1-10 (2024). <https://doi.org/10.1088/1755-1315/1315/1/012077>
10. J.Y. Lee, M. Kong, J. Oh, J.S. Lim, S.H. Chung, J.M. Kim, J.S. Kim, K.H. Kim, J.C. Yoo, W. Kwak, Comparative evaluation of Nanopore polishing tools for microbial genome assembly and polishing strategies for downstream analysis. *Sci Rep*. **11**, 1-11 (2021). <https://doi.org/10.1038/s41598-021-00178-w>
11. H. Zhiqi, W. Tingyi, C. Dongdong, S. Lan, Z. Guangheng, Q. Qian, Z. Li, Leucine-Rich Repeat Protein Family Regulates Stress Tolerance and Development in Plants. *Rice Sci*. **32**, 32–43 (2025). <https://doi.org.10.1016/j.rsci.2024.12.003>
12. S.B. Janssens, F. Vandeloek, E. De Langhe, B. Verstraete, E. Smets, I. Vandenhouwe, R. Swennen. Evolutionary dynamics and biogeography of *Musaceae* reveal a correlation between the diversification of the banana family and the geological and climatic history of Southeast Asia. *New Phytologist*. **210**, 1453–1465 (2016). <https://doi.org/10.1111/nph.13856>

13. Y. Saijo, E.P. Loo iian, S. Yasuda. Pattern recognition receptors and signaling in plant–microbe interactions. *Plant Journal*. **93**, 592–613 (2018).
<https://doi.org/10.1111/tpj.13808>
14. S. Wang, S. Han, X. Zhou, C. Zhao, L. Guo, J. Zhang, F. Liu, Q. Huo, W. Zhao, Z. Guo. Phosphorylation and ubiquitination of OsWRKY31 are integral to OsMKK10-2-mediated defense responses in rice. *Plant Cell*. **35**, 2391–2412 (2023).
<https://doi.org/10.1093/plcell/koad064>
15. X.Q. Yu, H.Q. Niu, C. Liu, H.L. Wang, W. Yin, X. Xia. PTI-ETI synergistic signal mechanisms in plant immunity. *Plant Biotechnol J*. **22**, 2113–2128 (2024).
<https://doi.org/10.1111/pbi.14332>