genes

MDPI

Article

Genome-Wide In-Depth Identification of NLR and PRR Immune Receptor Genes in Mango and Their Population Variation and Pathogen-Responsive Expression

Baihui Jin [1,†], Jian Dong [2,†], Maoren Bai [1], Bisheng Shi [1], Mingyang Gan [1], Kaiya Sun [1], Yusong Zhang [3,*] and Xing Zhang [4,*]

[1] Faculty of Agronomy and Life Science, Kunming University, Kunming 650201, China; baimaoren@126.com (M.B.)
[2] College of Landscape and Horticulture, Yunnan Agricultural University, Kunming 650201, China
[3] Industrial Crops Research Institute, Yunnan Academy of Agricultural Sciences, Kunming 650205, China
[4] College of Landscape and Horticulture, Yunnan Forestry Technological College, Kunming 650224, China
* Correspondence: zhangys@yaas.org.cn (Y.Z.); zhangxingyn@126.com (X.Z.)
† These authors contributed equally to this work.

Abstract

Background/Objectives: Mango (*Mangifera indica* L.) is threatened by *Colletotrichum gloeosporioides* anthracnose. We examined the genomic organization, population variation and infection responses of nucleotide-binding leucine-rich repeat (NLR) and pattern-recognition receptor (PRR) genes in mango. **Methods:** We identified NLR and PRR genes in CATAS_Mindica_2.1 with Resistify and SignalP 6.0 and surveyed 222 resequenced accessions. After filtering (≥80% call rate, biallelic status and minor allele frequency ≥5%), 742 NLR and 2149 PRR sites were retained for principal component analysis (PCA), nucleotide diversity and Hudson FST. We analysed 36 infected-fruit RNA-seq libraries, fungal load, W-box elements and small secreted-protein candidates (<400 aa). **Results:** We identified 592 NLR genes (1879 transcripts), including 134 with a canonical CC-NB-ARC-LRR architecture, and 838 PRR genes (1049 transcripts). Mean call rates were 91.9% and 92.4%. PCA and population differentiation (FST) showed no geographic structure among populations with n ≥ 3 (PC1/PC2 9.9%/5.4% for NLR and 7.8%/5.4% for PRR). Nucleotide diversity was higher in PRR than NLR genes (3.0 versus 15.4×10^{-5} per bp). Of 51 genes responsive in ≥2 contrasts, 45 were PRRs; NLRs were under-represented and PRRs enriched at 2 days post-inoculation (odds ratio 2.32, adjusted $p = 4.4 \times 10^{-4}$). PRR promoters were enriched in W-box elements (adjusted $p = 4.2 \times 10^{-8}$), but W-box copy number did not correlate with response magnitude ($\rho = 0.09$ and 0.08). **Conclusions:** Detectable PRR genes were more transcriptionally responsive than NLR genes, suggesting regulatory differences rather than a functional hierarchy. These results provide a genome-wide resource and testable hypotheses, although population analysis remains genotype-limited.

Keywords: Mangifera indica; anthracnose; Resistify; nucleotide diversity; plant immunity; effector

Check for updates

Academic Editors: Aijun Wang and Desuo Yin

Received: 7 November 2026
Revised: 23 November 2026
Accepted: 24 November 2026
Published: 26 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Mango (*M. indica* L.) is one of the most economically and culturally significant fruit crops cultivated across the tropical and subtropical regions of the world. Its sustainability and yield stability are threatened by a broad spectrum of phytopathogens throughout the entire developmental cycle. Among these, mango anthracnose, caused primarily by

Colletotrichum species and in particular by *C. gloeosporioides*, is a destructive disease that damages foliage, inflorescences and fruit, and it remains a major constraint on the global mango industry. Elucidating the genetic architecture of the mango innate immune system is therefore a prerequisite for durable resistance breeding.

Plants have evolved a sophisticated two-tiered immune system to perceive and counteract microbial intrusion. The first line of defense is mediated by pattern recognition receptors (PRRs) located at the plasma membrane. PRRs detect conserved pathogen- or microbe-associated molecular patterns and initiate pattern-triggered immunity, which restricts basal microbial colonization [1]. Pathogens that overcome this layer deploy effectors into host cells; in response, plants use intracellular nucleotide-binding leucine-rich repeat (NLR) proteins to recognize these effectors and trigger effector-triggered immunity, which frequently culminates in a localized hypersensitive response [2]. Because PRRs act at the front line of cell-surface surveillance while NLRs dictate intracellular, highly specific recognition, the two families are expected to experience different selective regimes and, consequently, different levels and patterns of nucleotide variation.

Genome-wide inventories of NLR and PRR genes are now available for several crops, and they consistently reveal that NLR genes are frequently clustered, copy-number variable and subject to diversifying selection, whereas PRR genes are more evenly distributed and more conserved. In mango, a comparative analysis of CC-NBS-LRR (CNL) genes in three cultivars identified 47, 27 and 36 CNL genes in Alphonso, Hong Xiang Ya and Tommy Atkins, respectively, and examined their stress-responsive expression [3]. That study was restricted to the CNL subclass and did not survey PRRs or population variation. Genome-wide immune-receptor repertoires have since been reported in other crops [4–6], and methods for annotation independent of transcript models [7] and for integrated-domain NLR architectures [8] have improved. Nevertheless, a unified inventory of NLR and PRR genes in a single mango reference genome, together with a population-level survey, is still lacking. High-quality reference genomes and large resequencing resources for mango have recently become available [9,10], providing an opportunity to place immune-receptor variation in a population context.

In this study we (i) performed a genome-wide identification and classification of NLR and PRR genes in the mango reference genome CATAS_Mindica_2.1 using Resistify; (ii) characterized single-nucleotide and short insertion/deletion variation within these genes across 222 globally distributed accessions; (iii) compared population structure, phylogenetic relationships, nucleotide diversity (π) and population differentiation (FST) between the two families using a strictly filtered genotype matrix; and (iv) used a publicly available *C. gloeosporioides* infection transcriptome of 36 libraries to test whether NLR and PRR genes are transcriptionally engaged during infection, and to quantify pathogen colonization and secreted-protein expression independently of the host. Because cell-surface and intracellular receptors share signaling components and can mutually potentiate immunity [11–15], we interpret the two families as interacting layers rather than as independent systems.

2. Materials and Methods

2.1. Data Sources

The mango reference genome and annotation (CATAS_Mindica_2.1, cultivar Alphonso; [9]) were used throughout; the 20 pseudochromosomes are covered by the accessions NC_058137.1–NC_058156.1 (additional unplaced scaffolds are named Scaffold1, Scaffold2, …). Whole-genome resequencing data of 222 mango accessions were obtained from the Genome Sequence Archive under accession CRA016056, which comprises 224 sequenced accessions (CRR1117458–CRR1117681); all population-level analyses in this manuscript were performed on the 222 accessions that were present in the merged variant set (Tables S1 and S2). The two

accessions absent from the merged set (CRR1117651 and CRR1117657) were retained in the sequence archive description but not analyzed further.

For the transcriptional analysis, we used the public RNA-seq experiment PRJNA575336 [16], in which mango fruit were inoculated with *C. gloeosporioides* and sampled at 0, 2 and 7 days post-inoculation (dpi). The experiment contains four fruit/side combinations, red fruit red side (RFRS), red fruit green side (RFGS), green fruit red side (GFRS) and green fruit green side (GFGS), with three biological replicates each (36 single-end libraries, 61 bp, Illumina HiSeq 2500,Inc., San Diego, CA, USA). The *C. gloeosporioides* Nara gc5 reference genome and annotation (GCA_000319635.1; 55.6 Mb) were used to quantify fungal transcription.

2.2. Identification and Classification of NLR and PRR Genes

Protein sequences of all 52,078 annotated transcripts (32,071 protein-coding genes) were screened with Resistify v1.1.1 using default parameters [17]. Resistify combines domain identification with a machine-learning classifier and reports both a family-level class and a domain-architecture string for each transcript. A gene was assigned to a family when at least one transcript was classified accordingly, and genes whose transcripts carried different combinations of domains were recorded with all observed classes. Throughout, we distinguish the Resistify CNL class (332 genes) from the canonical CNL architecture (134 genes whose longest transcript contained coiled-coil, NB-ARC and LRR domains). Transmembrane and signal-peptide prediction for PRR genes was performed with SignalP 6.0 in slow (protein language model) mode [18]; kinase domains were taken from the Resistify domain output. Alternative transcripts were retained in the inventory, and the longest transcript per gene was used for gene-level architecture summaries. Family phylogenies were reconstructed from full-length protein sequences (not isolated domains) by MAFFT alignment [19] and FastTree inference [20].

2.3. Variant Identification and Functional Annotation

Raw reads of the 222 accessions were quality-filtered with fastp v0.23.4 [21] (-q 25), aligned to CATAS_Mindica_2.1 with BWA-MEM v0.7.17-r1188 [22] and processed with SAMtools v1.18 [23]. Variants were called per sample with DeepVariant v1.6.0 [24] and merged with bcftools v1.17 [25]. Variants located within NLR or PRR gene intervals (obtained from the Resistify annotation) were extracted and functionally annotated with ANNOVAR (version 20191024) [26] using a refGene database built from the CATAS_Mindica_2.1 GFF3 with gff3ToGenePred and retrieve_seq_from_fasta.pl.

Raw reads of the 222 accessions were quality-filtered with fastp v0.23.4 [21] (-q 25), aligned to CATAS_Mindica_2.1 with BWA-MEM v0.7.17-r1188 [22] and processed with SAMtools v1.18 [23]. Variants were called per sample with DeepVariant v1.6.0 [24] and merged with bcftools v1.17 [25]. Variants within NLR or PRR gene intervals were extracted and annotated with ANNOVAR [26] using a refGene database built from the CATAS_Mindica_2.1 GFF3. The merged call set was dominated by missing genotypes: mean per-site call rates were 4.04% for NLR and 8.34% for PRR sites in the complete extracted sets (Table S7). To avoid basing population analyses on this matrix, we retained only sites with a call rate ≥80%, exactly two alleles, and a minor allele frequency (MAF) ≥5%. This yielded 742 NLR and 2149 PRR sites in the PHYLIP alignments used for genotype-based analyses; the corresponding bcftools-filtered VCF contained 807 and 2305 sites, the difference reflecting sites that could not be represented as strict biallelic IUPAC codes. Missing genotypes were never converted to homozygous reference. For PCA only, missing genotypes were replaced by the site allele frequency (mean imputation); all other analyses used complete-case genotypes at each site. Per-sample call rates at the retained sites were 91.9% ± 4.2% (NLR) and 92.4% ± 3.8% (PRR) (mean ± SD).

2.4. Population Structure and Phylogenetic Analysis

The filtered genotype matrices were analyzed with PCA (222 accessions × 742 or 2149 sites) after mean imputation, and the first two components were visualized. For phylogenetic reconstruction, the filtered alignments were converted to PHYLIP [27] and trees were inferred with IQ-TREE v2.2.0.3 using GTR+F+G4+ASC, an ascertainment-bias-corrected GTR model with four rate categories, as recorded in the IQ-TREE log [28]. Trees were annotated with itol.toolkit v1.1.7 [29] and displayed with iTOL [30]. Split networks were computed with SplitsTree [31] from the same alignments and are shown as exploratory visualizations only.

2.5. Nucleotide Diversity and Population Differentiation

Nucleotide diversity (π) was estimated from the filtered genotypes as the average number of pairwise differences per bp, using the total length of the NLR or PRR gene intervals as the denominator; 95% confidence intervals were obtained by bootstrapping sites (5000 replicates). Because the retained sites are MAF-filtered, these estimates are not directly comparable to unfiltered genome-wide π; we therefore emphasize the direction of the family contrast rather than its absolute magnitude. As a sensitivity analysis, we also report VCFtools v0.1.16 [32] window π (100-kb windows, 1-kb step) calculated from the complete extracted sets. Pairwise differentiation was estimated with Hudson's ratio-of-averages estimator from the filtered sites. Only the 10 populations represented by at least three accessions were compared (45 pairs); Mexico, Brazil and Sri Lanka ($n = 1$) and Israel ($n = 2$) were excluded from statistical comparisons and are reported descriptively only. Negative *F*ST estimates were retained and interpreted as absence of detectable differentiation, not as negative differentiation.

2.6. RNA-Seq Data Processing and Differential Expression Analysis

Raw reads of the 36 libraries were downloaded from the CNCB mirror of the INSDC archive, converted with fasterq-dump (SRA Toolkit), trimmed with fastp (Phred quality 15, minimum length 30 nt) and quantified against a transcriptome extracted from the CATAS_Mindica_2.1 annotation with gffread v0.12.8 [33], yielding 63,442 transcripts, of which 52,078 corresponded to protein-coding mRNAs. Quantification was performed with Salmon v0.13.1 [34] in selective-alignment mode (–validateMappings, k-mer size 23, automatic library-type detection); transcript-level estimates were collapsed to the gene level by summing NumReads and TPM.

Raw reads of the 36 libraries were downloaded from the CNCB mirror of the INSDC archive, converted with fasterq-dump (SRA Toolkit), trimmed with fastp (Phred quality 15, minimum length 30 nt) and quantified against a transcriptome extracted from the CATAS_Mindica_2.1 annotation with gffread v0.12.8 [33], yielding 63,442 transcripts of which 52,078 corresponded to protein-coding mRNAs. Quantification was performed with Salmon v0.13.1 [34] in selective-alignment mode (–validateMappings, k-mer size 23, automatic library-type detection); transcript-level estimates were collapsed to the gene level by summing NumReads and TPM. Genes with at least 10 counts in at least three libraries were retained, leaving 19,548 of 32,071 genes for testing. Counts were modeled with DESeq2 [35] using a design that combined fruit/side combination and time point; eight contrasts were tested, corresponding to each combination at 2 and 7 dpi against its own 0 dpi control. Log2 fold changes were shrunken with the normal prior, and genes with a Benjamini–Hochberg adjusted $p < 0.05$ and $|\log2 \text{ fold change}| \geq 1$ were considered differentially expressed (DE). Variance-stabilized counts were used for PCA and sample distance heat maps. All family-level enrichment tests used the 19,548 testable genes as

the background, so that differences in the number of testable NLR and PRR genes were explicitly accounted for.

2.7. Estimation of Fungal Load

To quantify fungal colonization independently of host transcription, trimmed reads were aligned to the *C. gloeosporioides* Nara gc5 genome. Three aligners were benchmarked on 300,000 reads: minibwa v0.7-r424 [36] aligned 11.51% of reads in 0.53 s with 461 MB peak memory, BWA-MEM v0.7.17 [22] aligned 11.49% in 1.63 s with 519 MB, and Bowtie2 v2.5.4 [37] aligned 4.77% in 1.32 s with 189 MB (Bowtie2 pairwise end-to-end mode being more stringent). minibwa was selected because it matched BWA-MEM in sensitivity while being the fastest of the three. Secondary and supplementary alignments were discarded, and fungal load was defined as the proportion of quality-filtered reads aligned with MAPQ $\geq$ 20; a stricter MAPQ $\geq$ 30 threshold gave equivalent results. This is a sequencing-derived proxy for fungal RNA abundance, not a direct measurement of biomass; ambiguous host–fungal alignments were addressed by the MAPQ threshold and by confirming that the host mapping rate declined in proportion to fungal load.

2.8. Family-Level Response Analysis of NLR and PRR Genes

For each contrast, the proportion of DE genes within the NLR and PRR families was compared with the genome-wide background (19,548 testable genes) using Fisher's exact test with Benjamini–Hochberg correction across all family–contrast combinations. Genes that were DE in at least two contrasts were designated core responsive genes. For every NLR and PRR gene, the Pearson correlation between log2(TPM + 1) and fungal load was computed across the 36 libraries and corrected for multiple testing. To test whether red and green fruit differ in their transcriptional response, an interaction model (~fruit + day + fruit:day) was fitted with DESeq2, complemented by a per-gene linear model of expression on fungal load, fruit type and their interaction. For the promoter analysis, the Spearman correlation between W-box copy number and the maximum absolute log2 fold change across the eight contrasts was computed separately for testable NLR and PRR genes.

2.9. Promoter Cis-Element Analysis

For each NLR and PRR gene, the 2000-bp sequence upstream of the annotated start codon was extracted from the reference genome, taking strand into account (1430 genes in total, all 2000 bp). Eight defense-related cis-elements (W-box TTGACC, W-box variant TTGACT, GCC-box GCCGCC, DRE/CRT CCGAC, G-box CACGTG, TGA/as-1 TGACG, MYB/MBS TAACTG and TC-rich TTTCTTCA) were counted on both strands, and copy numbers were compared between families (Wilcoxon rank-sum test) and between responsive and non-responsive genes (Wilcoxon and Fisher tests). W-box copy number was also correlated with the magnitude of the transcriptional response, as described above.

2.10. Prediction of Secreted Proteins in C. gloeosporioides

The 15,381 predicted proteins of *C. gloeosporioides* Nara gc5 were screened for secretion with SignalP 6.0 (eukaryotic mode) and DeepLoc 2.1 [38]. Proteins carrying a signal peptide and predicted to localize extracellularly were designated secreted-protein candidates. Because small secreted proteins are the class most often associated with pathogen virulence, this analysis was restricted to candidates shorter than 400 amino acids, yielding 1035 small secreted-protein candidates, not confirmed effectors. EffectorP 3.0 [12] and structural comparisons [13] were not run in this study; these 1035 candidates should therefore be regarded as a prioritized subset for future effector prediction. Fungal gene expression was quantified by aligning the same trimmed reads to the fungal genome with minibwa, retaining alignments with MAPQ $\geq$ 20 and

counting reads per gene with featureCounts [39] using the GCA_000319635.1 GFF3 annotation; counts were converted to CPM after library size normalization.

2.11. Software and Reproducibility

All command lines, versions and intermediate files are recorded in the project CHANGELOG.md and in the README.md files of the corresponding result/sub-directories. R analyses were run with R 4.6.1; genomic command-line analyses were run with bcftools 1.22, VCFtools 0.1.16, PLINK 1.9 and featureCounts 2.0.6 [40]. The filtered genotype matrices, retained-site lists and analysis scripts are provided as Supplementary Data so that the population analyses can be reproduced independently.

3. Results

3.1. Genome-Wide Identification of NLR Genes

Resistify identified 592 NLR genes in the mango genome, corresponding to 1879 transcripts and 1.85% of the 32,071 protein-coding genes (Figure 1; Table S4). The genes are unevenly distributed: 19 of the 20 pseudochromosomes carry at least one NLR gene (509 genes in total), and 18 unplaced scaffolds carry a further 83 genes. Chromosome 2 carries by far the largest number (111 genes), followed by chromosomes 6 (64), 7 (53) and 16 (48); chromosome 5 carries only a single NLR gene and chromosome 8 carries none (Table S3). Two scaffolds are particularly NLR-rich (Scaffold1, 19 genes; Scaffold2, 32 genes), indicating local clustering of immune receptors.

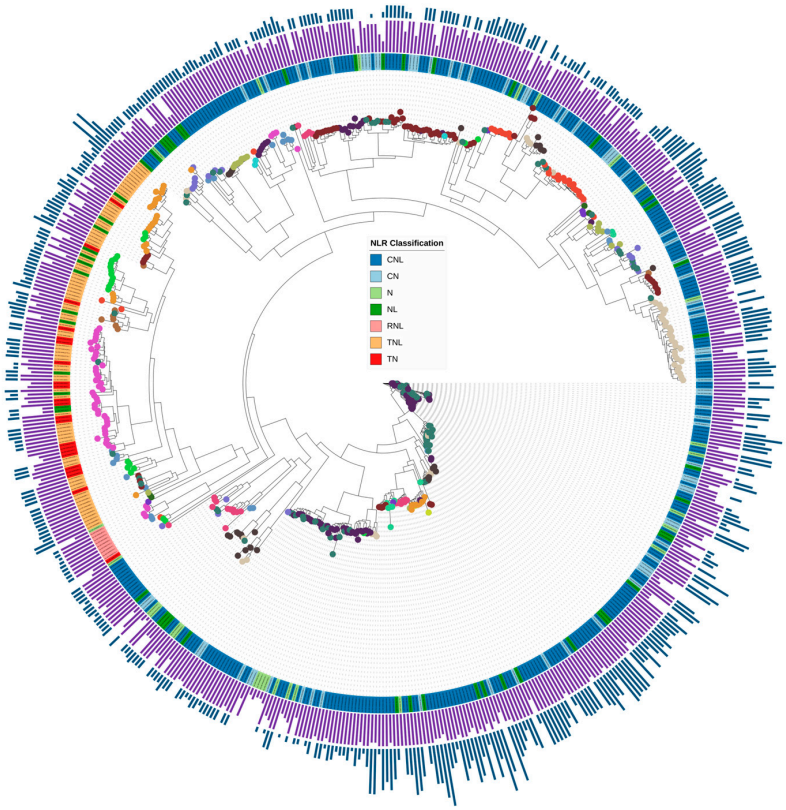


Figure 1. Phylogenetic relationships of NLR genes in the mango genome (CATAS_Mindica_2.1). The circular tree was inferred with FastTree from the MAFFT alignment of 592 full-length NLR protein sequences. Terminal points are colored by chromosome or scaffold. Individual gene labels are omitted at this scale; complete gene identifiers and structural annotations are provided in the Source Data sheet, Figure 1 and Table S4.

At the gene level, Resistify assigned 332 genes to the CNL class, 84 to CN, 79 to TNL, 56 to NL, 28 to N, 24 to TN and 9 to RNL, giving 612 class assignments for 592 genes (Table S4). This excess reflects 20 genes whose alternative transcripts fell into different classes. The domain-architecture strings on the longest transcript were more heterogeneous than the class labels: 134 genes had a canonical CNL architecture, followed by NL (137), N (64), NLL (59), TNLj (45), TNL (33), CN (31), mCNL (21), TN (20), NLLL (19), RNL (9) and 20 genes with other combinations (Table S4a,b; Source Data sheet Table S4_domain_architecture). We therefore use "CNL class" for the Resistify assignment and "canonical CNL architecture" for the 134 genes with CC, NB-ARC and LRR domains; the two counts should not be used interchangeably. Protein length of the 1879 transcripts ranged from 90 to 4211 amino acids, the LRR domain from 0 to 1982 amino acids (mean 510.7) and the number of NB-ARC motifs from 1 to 9. Very short predictions were retained in the inventory but should be treated as partial models; they may reflect incomplete annotation or alternative transcripts rather than genuine short receptors.

3.2. Genome-Wide Identification of PRR Genes

The same procedure identified 838 PRR genes (1049 transcripts), accounting for 2.61% of the protein-coding genes (Figure 2; Table S5). In contrast to the NLR family, PRR genes are present on all 20 pseudochromosomes and on 30 scaffolds, and their distribution is far more even: chromosome 1 carries the largest number (105 genes) and chromosome 19 the smallest (17 genes), with the remaining chromosomes carrying between 20 and 61 genes. Notably, chromosome 2, which harbours the greatest number of NLR genes, carries only 33 PRR genes, whereas chromosome 5, with a single NLR gene, carries 34 (Table S3).

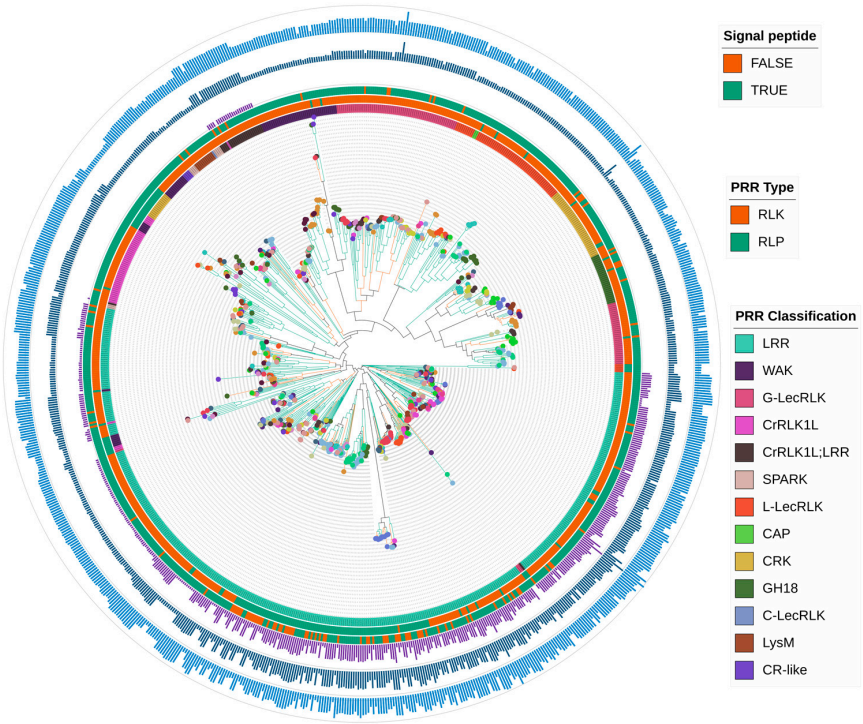


Figure 2. Phylogenetic relationships of PRR genes in the mango genome. The circular tree was inferred with FastTree from the MAFFT alignment of 838 full-length PRR protein sequences. Terminal points are colored by chromosome or scaffold and shaped according to receptor type (RLK or RLP). Individual gene labels are omitted at this scale; complete gene identifiers and annotations are provided in Source Data sheet, Figure 2 and Table S5.

At the gene level, 673 PRR genes encode receptor-like kinases and 165 encode receptor-like proteins. Thirteen structural classes were recovered, of which the LRR class was the most abundant (431 genes), followed by G-LecRLK (98), L-LecRLK (65), WAK (65), CRK (52), CrRLK1L (48), GH18 (26), CrRLK1L/LRR (25), LysM (11), SPARK (10) and the rare classes CR-like (4), CAP (2) and C-LecRLK (1). SignalP 6.0 predicted a signal peptide for 735 of the 838 PRR proteins, consistent with a cell-surface tendency, but a signal peptide alone does not establish plasma-membrane localization or immune function. These 838 genes are therefore computational PRR candidates; not every RLK or RLP is a pathogen-recognition receptor, and functional annotation will require expression, localization and ligand studies.

3.3. Variation Within NLR and PRR Genes Across 222 Accessions

Resequencing data of 222 mango accessions were aligned to CATAS_Mindica_2.1 and variants were called and annotated. A total of 147,199 single-nucleotide polymorphisms (SNPs) and 1818 InDels were detected within NLR gene intervals, and 101,605 SNPs and 855 InDels within PRR gene intervals (Table S6). ANNOVAR assigned 42,808 NLR SNPs (29.1% of all NLR SNPs) and 46,403 PRR SNPs (45.7% of all PRR SNPs) to exons; the remaining SNPs were intronic (56.0% and 42.4%), in untranslated regions (8.2% and 11.0%), at splice sites (0.2% and 0.2%), intergenic (4.7% and 0.2%), in ncRNA regions (1.2% and 0.4%) or upstream/downstream of genes (0.6% and 0.2%) (Table 1). These first-level categories are mutually exclusive; Table 1 gives two-decimal percentages that sum to 100.00% within each family, and the nonsynonymous and synonymous percentages below are subcategories of the exonic class that must not be added to the first-level percentages.

Table 1. First-level ANNOVAR categories for SNPs within NLR and PRR gene intervals. Categories are mutually exclusive and sum to 100.00% within each family. Nonsynonymous and synonymous counts are subcategories of the exonic class and are reported in the text and Table S6; underlying data are in Source Data sheet, Table 1_ANNOVAR.

Family	ANNOVAR Category	SNPs (*n*)	Percentage (%)
NLR	Exonic	42,808	29.08
NLR	Intronic	82,458	56.02
NLR	UTR	11,997	8.15
NLR	Intergenic	6977	4.74
NLR	ncRNA	1790	1.22
NLR	Upstream/downstream	858	0.58
NLR	Splice site	311	0.21
NLR	Total	147,199	100.00
PRR	Exonic	46,403	45.67
PRR	Intronic	43,035	42.36
PRR	UTR	11,179	11.00
PRR	Intergenic	183	0.18
PRR	ncRNA	448	0.44
PRR	Upstream/downstream	176	0.17
PRR	Splice site	181	0.18
PRR	Total	101,605	100.00

The exonic variants were dominated by amino-acid-changing substitutions. In NLR genes, 28,993 nonsynonymous SNVs (67.7% of exonic variants) and 12,651 synonymous SNVs (29.6%)

were found, together with 1077 stopgain and 43 stop-loss variants and 44 variants of unknown consequence. In PRR genes, 27,820 nonsynonymous SNVs (60.0%) and 17,685 synonymous SNVs (38.1%) were detected, with 867 stopgain and 31 stoploss variants (Table S6a,b, Figure 3B,C). Thus, most of the exonic variation segregating in both families alters the encoded protein, and the ratio of nonsynonymous to synonymous exonic variants is slightly higher for NLR (2.29) than for PRR (1.57) genes. These ratios describe the variant spectrum and are not a formal test of selection.

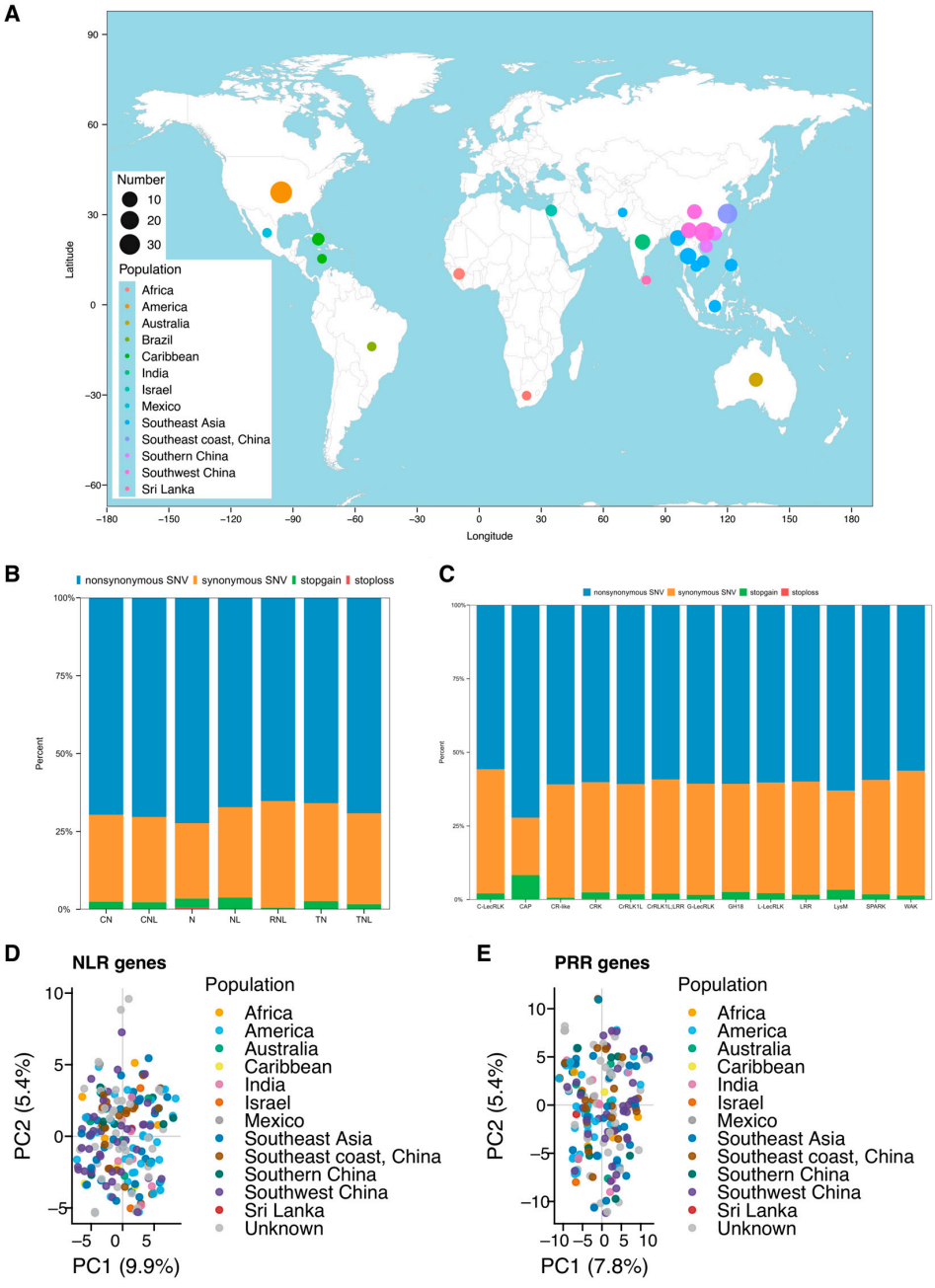


Figure 3. Geographic origin, exonic variant spectra and population structure of the 222 mango accessions. (**A**) Geographic distribution; circle size is proportional to the number of accessions and color denotes population. (**B**,**C**) Exonic variant spectra within NLR (**B**) and PRR (**C**) genes. (**D**,**E**) PCA of the filtered genotype matrices (742 NLR and 2149 PRR sites); variance explained is shown on each axis. Data: Tables S1, S2 and S6; Source Data Figure 3_PCA_NLR and Figure 3_PCA_PRR.

3.4. Population Structure and Phylogenetic Relationships at Immune Receptor Loci

The 222 accessions originate from 14 geographical regions spanning Africa, America, Australia, the Caribbean, India, Israel, Mexico, Sri Lanka and East and Southeast Asia, with the largest groups being Southwest China (42 accessions), America (37), Southeast Asia and unassigned accessions (34 each), and Southeast coast of China (26) (Tables S1 and S2, Figure 3A).

After strict filtering, 742 NLR and 2149 PRR sites were retained, with mean per-site call rates of 91.9% and 92.4% and per-sample call rates of 91.9% ± 4.2% and 92.4% ± 3.8% (mean ± SD). PCA of the filtered NLR matrix placed 9.9% of the variance on PC1 and 5.4% on PC2; the corresponding values for the PRR matrix were 7.8% and 5.4% (Figure 3D,E). In both analyses the accessions formed a single diffuse cloud without separation of the geographical groups, and no individual was a clear outlier. The maximum-likelihood trees were poorly resolved, with many zero-length branches and no geographically coherent clades (Figure S6A,B); the split networks were extensively reticulate (Figure S6C,D). We interpret the trees and networks as descriptive visualizations: reticulation is not, by itself, evidence of historical gene flow, and the data do not identify the demographic process behind the absence of structure.

3.5. Nucleotide Diversity Is Higher in PRR than in NLR Genes

Nucleotide diversity was higher in PRR than in NLR genes in the high-quality subset: 3.0×10^{-5} per bp (95% bootstrap CI 2.9–3.1 $\times 10^{-5}$) for NLR genes and 15.4×10^{-5} per bp (15.2–15.7 $\times 10^{-5}$) for PRR genes, a 5.1-fold difference (Figure 4A). The same direction was observed in the full window-based analysis (mean $\pi = 1.80 \times 10^{-5}$ over 31,871 NLR windows versus 2.64×10^{-5} over 53,905 PRR windows; Wilcoxon rank-sum test, $p < 2.2 \times 10^{-16}$; Figure S8; Table S8c). The absolute values differ between the two analyses because the high-quality subset is MAF-filtered and because the window analysis uses a different denominator; we therefore interpret the family contrast, not the magnitude. Among the 10 populations with $n \geq 3$, per-population π ranged from 2.9 to 4.3×10^{-5} for NLR genes and from 1.5 to 2.0×10^{-4} for PRR genes (Figure 4B,C; Table S8a,b). Israel ($n = 2$), Mexico ($n = 1$), Brazil ($n = 1$) and Sri Lanka ($n = 1$) were excluded from statistical comparison; their estimates are shown only as descriptive values in Table S8 and are not interpreted as population parameters.

3.6. Population Differentiation Between Immune Receptor Genes Is Negligible

Pairwise *F*ST was estimated with Hudson's estimator from the filtered sites for the 10 populations with $n \geq 3$ (45 pairs). All 45 estimates were ≤0 (range −0.254 to −0.011; median −0.09), indicating no detectable differentiation at these immune-receptor loci (Table S9, Figure S7). Negative estimates are expected when true differentiation is near zero and sampling variance dominates; they should not be interpreted as negative differentiation or as evidence of homogenizing selection. No population pair showed appreciable genome-wide differentiation, in agreement with the PCA. This is a descriptive statement about the present data, not a reconstruction of mango demographic history.

3.7. Colletotrichum Colonization Diverges Strongly Between Green and Red Fruit

To relate immune-receptor variation to infection, we turned to an experimentally inoculated mango fruit system. Because the host mapping rate of heavily infected libraries was unusually low, we first quantified fungal colonization directly by aligning the same reads to the *C. gloeosporioides* genome. At 0 dpi, fungal load was nearly identical in all four fruit/side combinations (1.20–1.27% of quality-filtered reads), providing an internal baseline. By 7 dpi, fungal load had diverged by more than an order of magnitude: 18.46% ± 3.44 in green fruit red side, 7.38% ± 0.48 in green fruit green side, 3.85% ± 0.46 in red fruit green side and 1.39% ± 0.27 in red fruit red side (Figure 5A; Table S14). Averaged over the two inoculated sides, fungal load at 7 dpi was 4.9-fold higher in green fruit than in red fruit (12.92% versus 2.62%),

independently reproducing the greater susceptibility of green mango fruit described by [16]. Because this metric is a sequencing-derived proxy, we use it to rank colonization rather than as an absolute measure of fungal biomass.

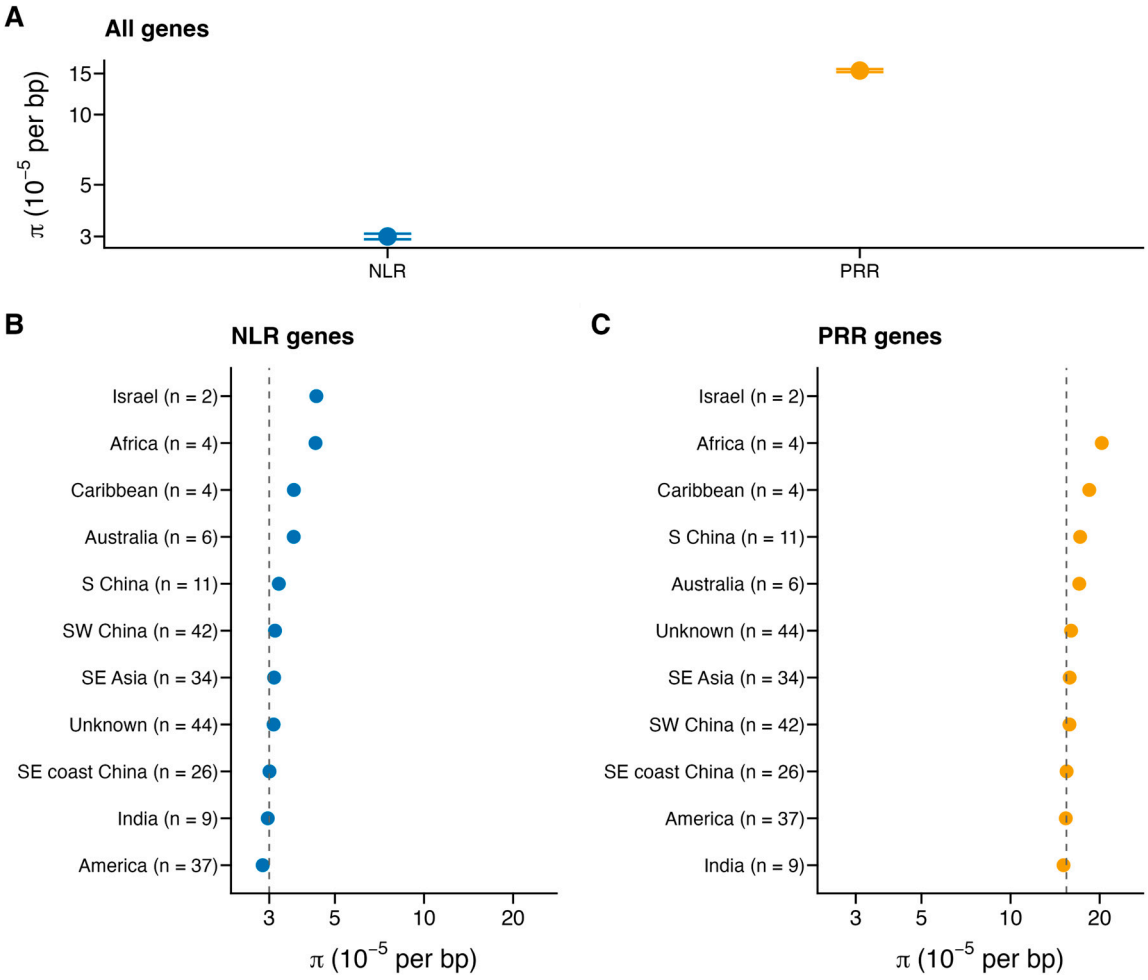


Figure 4. Nucleotide diversity (π) of NLR and PRR genes. (**A**) π per bp estimated from the high-quality subset (742 NLR and 2149 PRR sites); error bars are 95% bootstrap confidence intervals (5000 site resamples). (**B**,**C**) Population-wise π per bp for NLR (**B**) and PRR (**C**) genes; populations are ordered by median π and the number of accessions is given in parentheses. Only populations with $n \geq 2$ are shown; Israel ($n = 2$) is descriptive only, and Mexico, Brazil and Sri Lanka ($n = 1$) are omitted because population diversity cannot be estimated from a single diploid genome. The dashed line marks the family mean. Because the retained sites are MAF-filtered, absolute values are not directly comparable to unfiltered genome-wide estimates; the family contrast is the robust result. Underlying data are given in Table S8 and Source Data sheets Figure 4_family and Figure 4_population.

Fungal load was strongly and negatively correlated with the proportion of reads mapping to the mango transcriptome (Pearson $r = -0.76$, $p = 7.8 \times 10^{-8}$), indicating that the reduced host mapping rate of the most heavily infected libraries reflects the accumulation of fungal RNA rather than compromised library quality. Consistently, the number of differentially expressed genes per contrast increased with fungal load ($r = 0.76$, $p = 0.028$; Figure 5B).

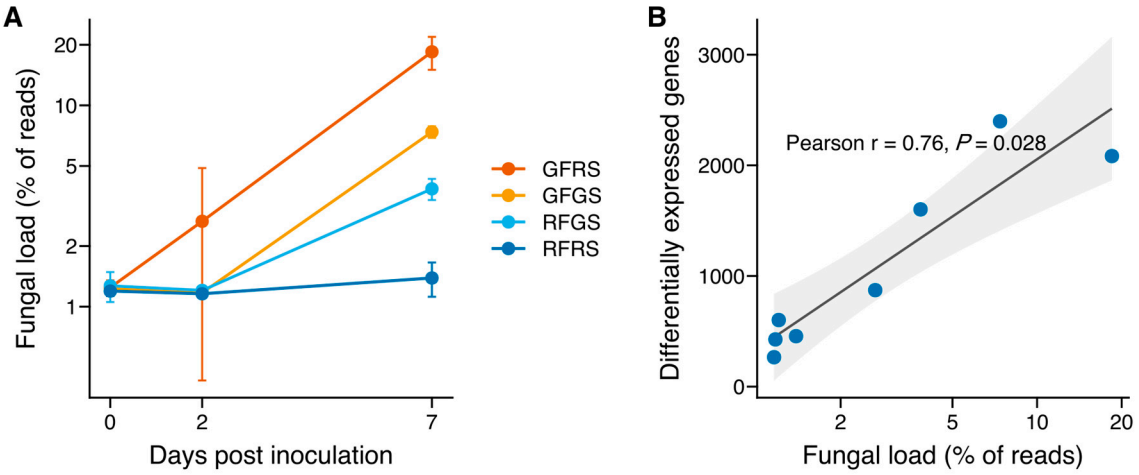


Figure 5. Fungal colonization of inoculated mango fruit. (**A**) Fungal load, defined as the percentage of quality-filtered reads aligned to the *C. gloeosporioides* Nara gc5 genome with MAPQ ≥ 20, in the four fruit/side combinations at 0, 2 and 7 days post-inoculation (dpi); points are means and error bars are SD ($n = 3$). (**B**) Relationship between the number of differentially expressed host genes and fungal load in the eight contrasts. Fungal load is a sequencing-derived proxy, not an absolute measure of biomass. Underlying data are given in Tables S11 and S14 and the Source Data sheets Figure 5_fungal_load and Figure 5_DE.

3.8. The Mango Transcriptome Responds Incrementally to Infection

Between 266 and 2398 genes were differentially expressed in the individual contrasts (Table S11). Responses were strongest at 7 dpi in green fruit (2398 DE genes in GFGS, 2084 in GFRS), intermediate in red fruit green side (1602) and weakest in red fruit red side (456), the combination with the lowest fungal load. PCA of variance-stabilized counts separated libraries primarily along PC1, which accounted for 25.1% of the total variance and followed the infection time course, with 0 dpi libraries at the negative extreme, 2 dpi libraries intermediate and 7 dpi libraries at the positive extreme (Figure 6A).

3.9. Detectable PRR Genes Are More Transcriptionally Responsive than NLR Genes

Of the 592 NLR and 838 PRR genes, 219 (37.0%) and 362 (43.2%), respectively, were expressed at levels sufficient for differential expression testing in fruit tissue (count filter: at least 10 counts in at least three libraries; Figure 7A; Table S12). Relative to the 19,548 testable genes, NLR genes were consistently under-represented among DE genes, with odds ratios between 0.10 and 0.45; the depletion was statistically significant at GFRS 2 dpi (OR = 0.10, Padj = 6.5×10^{-3}), GFRS 7 dpi (OR = 0.44, Padj = 0.017), GFGS 7 dpi (OR = 0.45, Padj = 0.010) and RFGS 7 dpi (OR = 0.21, Padj = 7.6×10^{-4}). PRR genes behaved in the opposite way and were significantly enriched among DE genes at GFRS 2 dpi (OR = 2.32, Padj = 4.4×10^{-4}). Because the background is restricted to testable genes, these tests compare the families with genes that passed the same expression filter; they do not imply that the non-testable members of either family are transcriptionally silent.

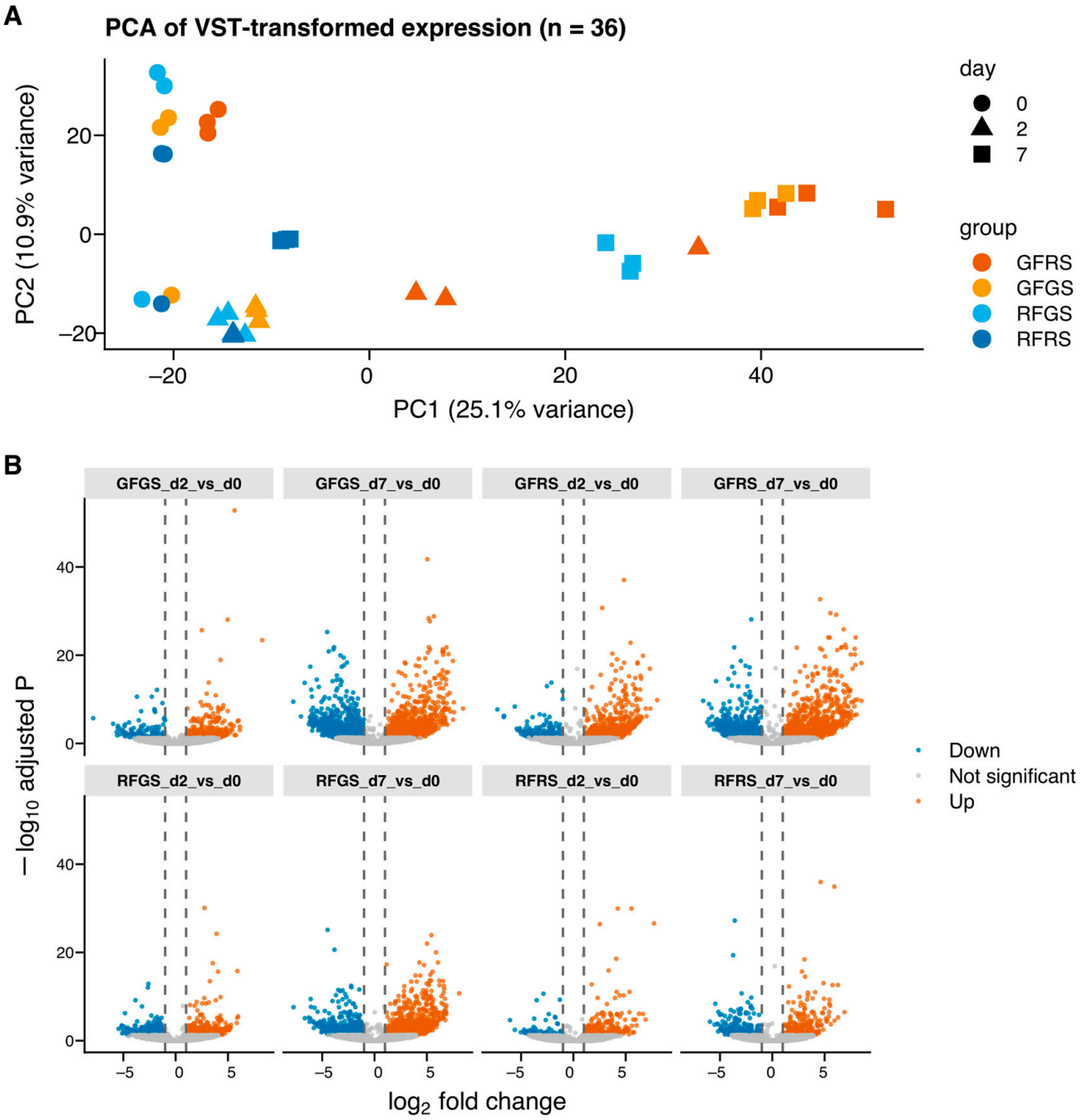


Figure 6. Transcriptome-wide response of mango fruit to *C. gloeosporioides* inoculation. (**A**) Principal component analysis of variance-stabilized gene counts; points are colored by fruit/side combination and shaped by time point. The percentage of variance explained is shown on each axis. (**B**) Volcano plots for the eight infection contrasts; points with adjusted $p < 0.05$ and |log2 fold change| ≥ 1 are shown as differentially expressed genes. Underlying data are given in the Source Data workbook (sheets Figure 6A,B).

The absolute number of responsive genes was highest for LRR-type PRRs (49 genes), followed by CNL-type NLRs (27), CRKs (26), G-LecRLKs (20), L-LecRLKs (20), CrRLK1Ls (16), WAKs (11) and LysM receptors (7) (Figure 7B). The response was therefore concentrated in cell-surface receptor classes, and the CNL signal was driven by a small number of genes rather than by the family as a whole. Transcript abundance is only one level of regulation; NLR-mediated immunity can also be activated through conformational change, oligomerization and downstream signaling without large changes in transcript levels [41].

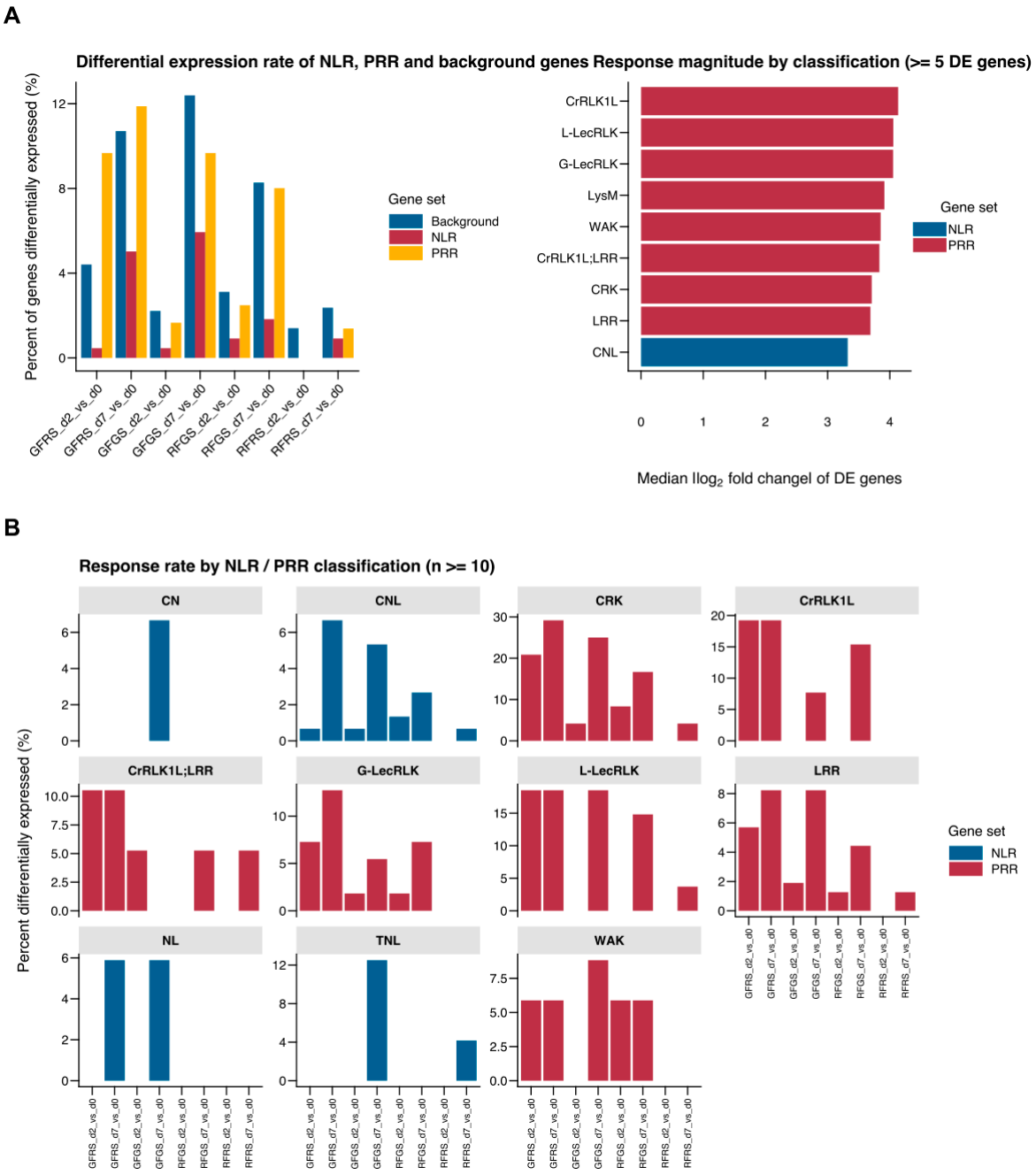


Figure 7. Differential expression of NLR and PRR genes during infection. (**A**) Proportion of differentially expressed genes in the NLR and PRR families compared with the genome-wide background for the eight contrasts; asterisks denote Fisher's exact tests that remain significant after Benjamini–Hochberg correction. (**B**) Number of differentially expressed genes per structural class. Underlying data are given in Table S12.

3.10. A Small Set of PRRs Constitutes the Transcriptional Core of the Response

Fifty-one NLR/PRR genes were differentially expressed in at least two contrasts; this core set was strongly biased towards cell-surface receptors, comprising 45 PRRs and only six NLRs (five CNL and one TNL) (Figure 8A; Table S13). The most broadly responsive gene was the CRK LOC123221601 (DE in seven of eight contrasts, maximum |log2 fold change| = 5.06), followed by the G-LecRLK LOC123223457 (five contrasts) and the CRKs LOC123218504 (four contrasts), together with additional CRK, CrRLK1L, L-LecRLK, LRR-type and LysM genes. Table 2 summarizes the 12 most broadly responsive candidates,

including direction, maximum effect size, adjusted *p* and correlation with fungal load; the complete set of 51 genes is provided in Table S13.

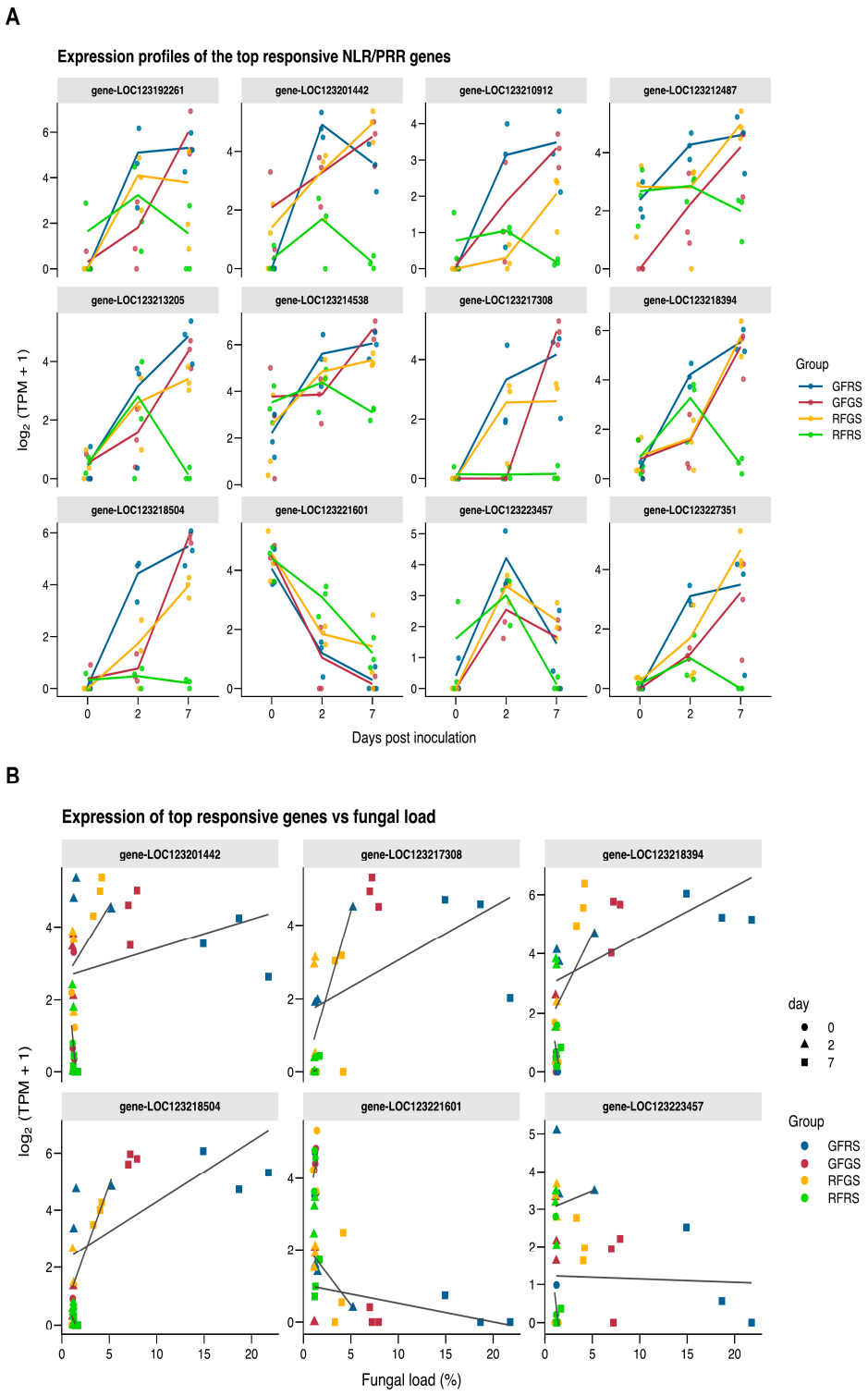


Figure 8. Core responsive NLR/PRR genes. (**A**) Expression trajectories of the 12 most broadly responsive genes across the four fruit/side combinations and three time points. (**B**) Correlations between expression and fungal load across all 36 libraries for six representative core genes. Complete data for all 51 core responsive genes are provided in Source Data sheets Figure 8A,B and Table S13.

Across all 36 libraries, 24 of the 51 core genes were significantly correlated with fungal load after multiple-testing correction: 17 PRRs were positively correlated and four PRRs

were negatively correlated, whereas all three correlated NLRs were negatively correlated (Figure 8B). Thus, PRR transcripts generally accumulated as the fungus proliferated, while the small subset of transcriptionally responsive NLRs declined. The strongest correlations were observed for the CRK LOC123218504 ($r = 0.71$, Padj $= 7.7 \times 10^{-5}$) and the CrRLK1L LOC123220699 ($r = -0.66$, Padj $= 1.6 \times 10^{-4}$). These correlations are descriptive and do not establish that fungal load causes the expression changes.

Table 2. The 12 most broadly responsive NLR/PRR candidate genes. The complete set of 51 core genes, with all contrasts and fungal-load correlations, is given in Table S13 and Source Data sheet Table 2_candidates.

Gene ID	Family	Class	DE Contrasts	Max |log2FC|	Min Adjusted p	Direction at 7 dpi	r with Fungal Load
LOC123221601	PRR	CRK	7/8	5.06	6.1×10^{-4}	down	−0.51
LOC123223457	PRR	G-LecRLK	5/8	4.26	2.1×10^{-2}	up	−0.10
LOC123218504	PRR	CRK	4/8	7.50	1.7×10^{-8}	up	+0.71
LOC123217308	PRR	LysM	4/8	6.65	4.4×10^{-5}	up	+0.61
LOC123201442	PRR	L-LecRLK	4/8	6.54	1.6×10^{-6}	up	+0.36
LOC123218394	PRR	CrRLK1L	4/8	6.48	7.8×10^{-9}	up	+0.63
LOC123210912	PRR	LRR	4/8	6.27	2.7×10^{-5}	up	+0.62
LOC123227351	PRR	L-LecRLK	4/8	5.90	8.2×10^{-5}	up	+0.40
LOC123213205	PRR	L-LecRLK	4/8	5.42	1.0×10^{-4}	up	+0.66
LOC123192261	PRR	CRK	4/8	5.34	4.6×10^{-4}	up	+0.55
LOC123214538	PRR	G-LecRLK	4/8	4.26	5.2×10^{-6}	up	+0.52
LOC123212487	PRR	LRR	3/8	6.64	3.6×10^{-8}	up	+0.45

3.11. PRR Promoters Are Enriched in W-Box Elements

To ask whether the difference in inducibility between the two families has a cis-regulatory correlate, we extracted the 2000-bp promoter regions of all 1430 NLR and PRR genes and counted eight defense-related elements on both strands. PRR promoters contained significantly more W-box elements, the binding site of WRKY transcription factors, than NLR promoters (mean 0.84 versus 0.55 copies per promoter, Wilcoxon Padj $= 4.2 \times 10^{-8}$), and the same was true for the TTGACT W-box variant (1.63 versus 1.22, Padj $= 1.7 \times 10^{-6}$) (Figure 9A; Table S15). The remaining six elements, including GCC-box, DRE/CRT, G-box, TGA/as-1, MYB/MBS and TC-rich elements, did not differ significantly between families after correction. Within each family, the copy number of none of the eight elements differed between genes that responded to infection and genes that did not (Padj > 0.49 for both NLR and PRR; Figure 9B). Moreover, W-box copy number did not correlate with the maximum absolute log2 fold change across the eight contrasts (NLR: $\rho = 0.09$, $p = 0.27$, $n = 141$; PRR: $\rho = 0.08$, $p = 0.23$, $n = 233$; Table 3 and Table S17a,b) or with the number of DE contrasts. The promoter difference is therefore a difference in predicted cis-regulatory architecture between the families, not evidence of a demonstrated WRKY-dependent mechanism, and it does not explain which individual receptor genes respond.

Table 3. Correlation between W-box copy number and the magnitude of the transcriptional response in testable NLR and PRR genes. Underlying data are in the Source Data sheet, Table 3_Wbox.

Family	Testable Genes (*n*)	Spearman Rho (W-Box vs. Max | log2FC|)	*p*	Adjusted *p*
NLR	141	0.094	0.267	0.267
PRR	233	0.078	0.233	0.267

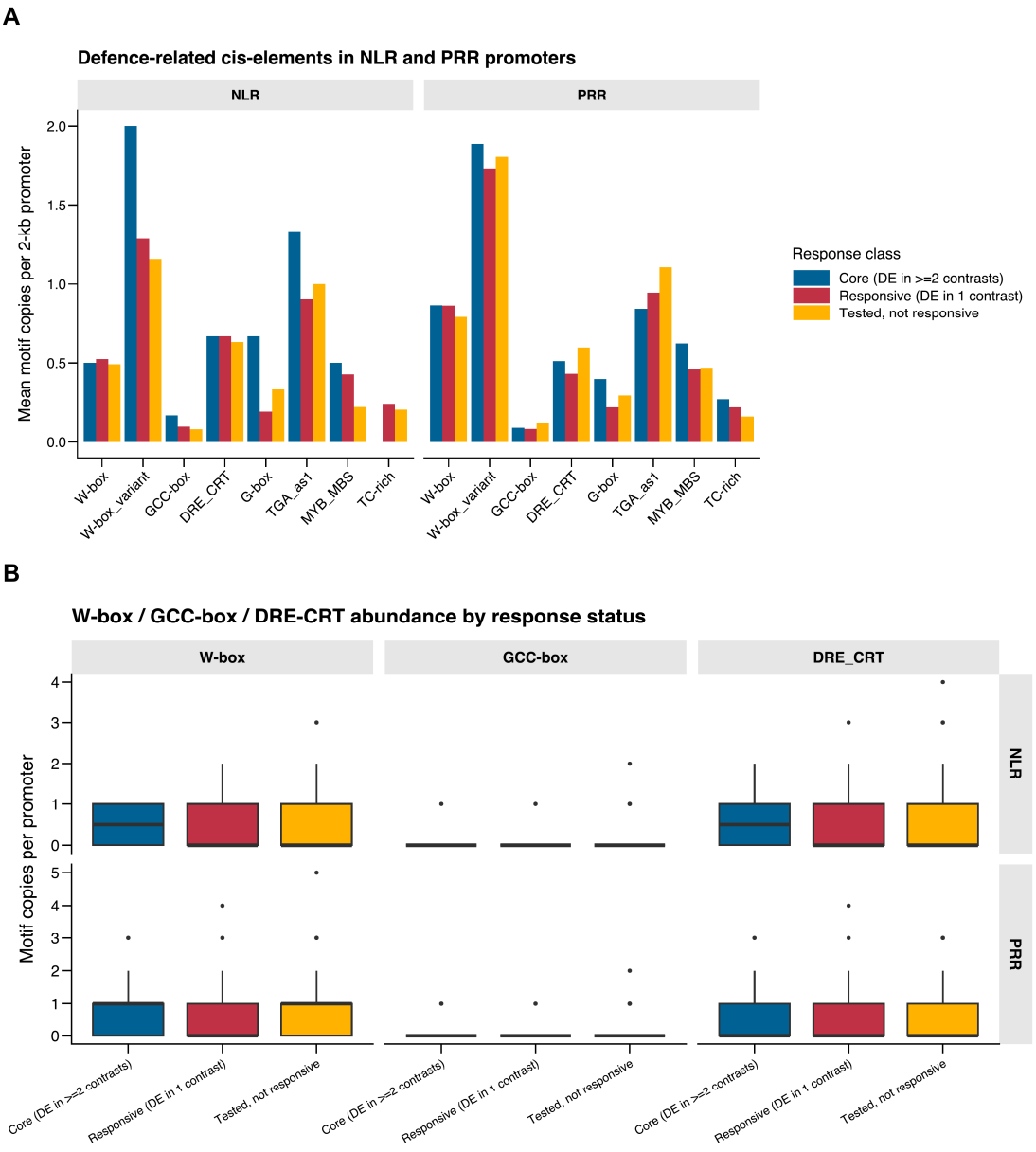


Figure 9. Cis-element architecture of NLR and PRR promoters. (**A**) Mean copy number of eight defense-related cis-elements in the 2000-bp promoter regions of NLR and PRR genes; asterisks indicate elements differing significantly between families after Benjamini–Hochberg correction. (**B**) Copy number of the same elements in genes that responded to infection versus genes that did not. Underlying data are given in Table S15.

3.12. The Pathogen Expresses Fewer Small Secreted Proteins in Resistant Fruit

To capture the pathogen side of the interaction, we predicted the secreted-protein repertoire of *C. gloeosporioides* Nara gc5 and quantified its expression in planta. Of the 15,381 predicted fungal proteins, 1903 carried a SignalP 6.0 signal peptide and 2142 were predicted

by DeepLoc 2.1 to localize extracellularly; the intersection defined 1717 secreted-protein candidates. Restricting this set to small secreted proteins (<400 aa) gave 1035 candidates (6.7% of the fungal gene set; Table S16). These are candidates, not functionally validated effectors; secretion and extracellular localization do not by themselves establish effector activity.

The share of fungal reads assigned to these small secreted proteins changed with infection: from 2.23% to 5.33% and 6.28% in GFRS, from 1.98% to 0.99% and 6.19% in GFGS, and from 1.32% to 2.18% and 6.66% in RFGS between 0, 2 and 7 dpi (Figure 10A). In the most resistant combination (red fruit, red side), the share remained low (0.77%, 1.89% and 1.44% at 0, 2 and 7 dpi), and the number of small secreted candidates induced at least four-fold at 7 dpi was also lowest in this combination (24), compared with 130 in GFRS, 88 in GFGS and 56 in RFGS. The most abundant candidate transcripts at 7 dpi were ELA28006 (CGGC5_1466; 3402 CPM), ELA24460 (CGGC5_14105; 2635 CPM) and ELA38278 (CGGC5_2675; 2240 CPM) (Figure 10B; Table S16). In resistant fruit, therefore, the fungus both accumulated less biomass and expressed a smaller secreted-protein program, complementing the host-side expression data. EffectorP 3.0 and structural comparisons are needed to determine which of these candidates are bona fide effectors [12,13].

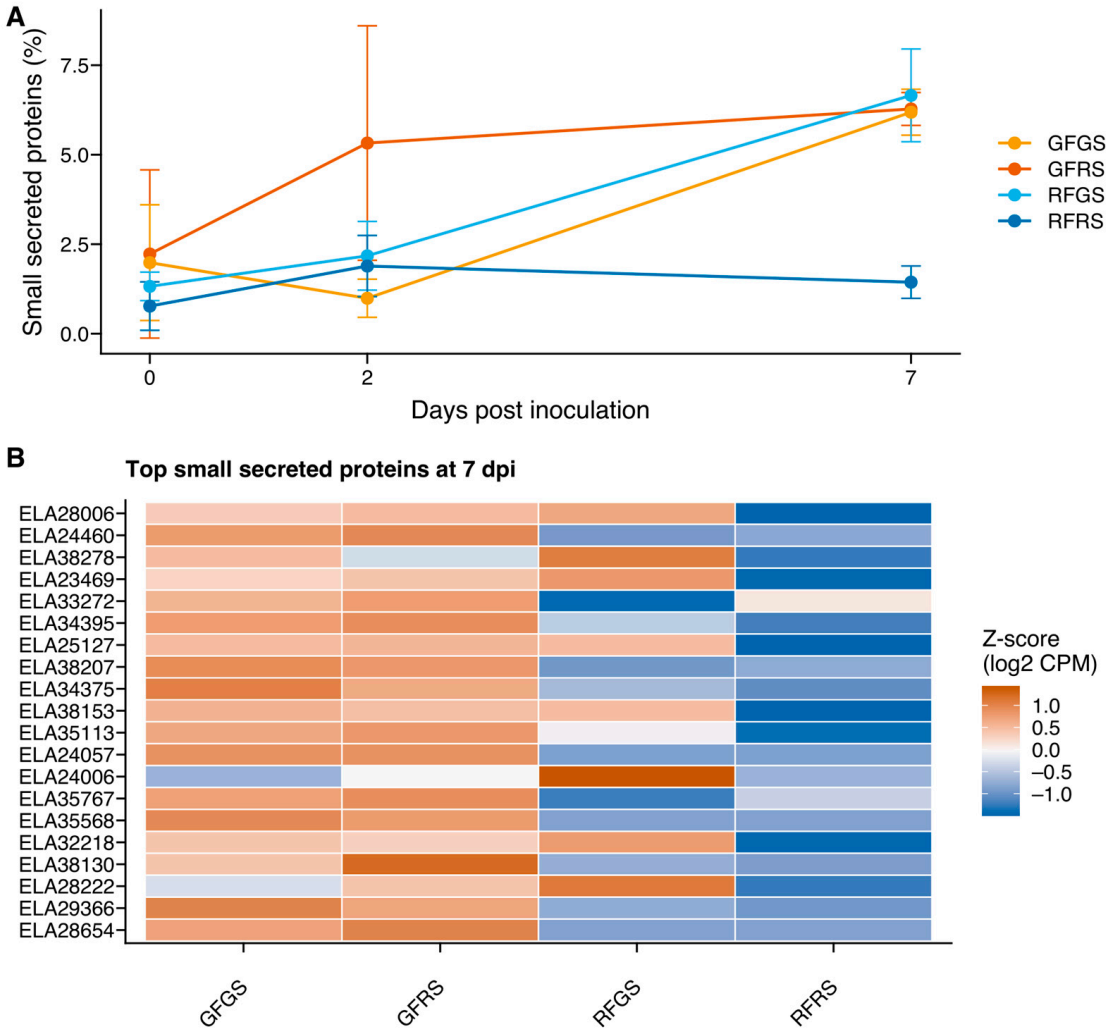


Figure 10. Expression of small secreted-protein candidates (<400 aa) of *C. gloeosporioides* during mango fruit infection. (**A**) Share of fungal reads assigned to the 1035 small secreted-protein candidates in each fruit/side combination over the infection time course. (**B**) Expression heat map of the 20 most abundant candidates at 7 dpi. These are computational candidates, not functionally validated effectors. Underlying data are given in Table S16.

4. Discussion

4.1. The Mango Immune Receptor Repertoire

We provide a genome-wide catalog of NLR and PRR genes in mango: 592 NLR genes (1879 transcripts) and 838 PRR genes (1049 transcripts), corresponding to 1.85% and 2.61% of the protein-coding genes. Both proportions lie within the range reported for other eudicot genomes, but the internal composition is informative. The NLR family is dominated by the CNL class (332 genes), the class most frequently implicated in effector recognition in dicots, and the family is strongly clustered: chromosome 2 alone carries 111 NLR genes and two unplaced scaffolds a further 51. Clustering of NLR genes is typical of plant genomes and is thought to favor the ectopic recombination and unequal crossing-over that generate new specificities. The PRR family, by contrast, is spread evenly over all 20 chromosomes and 30 scaffolds and is dominated by LRR-type receptors, with 673 receptor-like kinases and 165 receptor-like proteins; the presence of a signal peptide in 735 of the 838 proteins is consistent with, but does not prove, cell-surface localization.

The difference from the 47 CNL genes reported for Alphonso by Ul Qamar et al. [3] requires explanation. The two studies differ in query scope (CNL only versus all NLR and PRR genes), genome annotation (a three-cultivar pan-genome versus CATAS_Mindica_2.1), classification method (domain/HMM search versus Resistify), and counting unit (transcript versus gene). Within our data, the CNL class contains 332 genes but only 134 have a canonical CC-NB-ARC-LRR architecture on the longest transcript; the remaining CNL-class genes have NL, NLL, mCNL or other architectures. The 134 canonical CNL genes are the closest comparable number to the 47 reported previously, but the difference remains substantial and is likely to reflect the broader coiled-coil detection and the inclusion of partial or divergent models by Resistify. We therefore report both counts and the full architecture table (Table S4) and treat the absolute CNL number as method-dependent rather than as a biological constant. Table 4 summarizes the comparison with prior studies.

Table 4. Comparison of the present study with prior mango and plant immune-receptor studies. Underlying data are in Source Data sheet, Table 4_study_comparison.

Study	Species and Genome	Method	Receptor Scope	Genes Reported	Population/Expression
Ul Qamar et al., 2023 [3]	M. indica; Alphonso, Hong Xiang Ya and Tommy Atkins (pan-genome)	Domain/HMM search of CC-NBS-LRR genes	CNL only	47/27/36 CNL genes	No population analysis; stress transcriptomes
This study	M. indica; CATAS_Mindica_2.1 (Alphonso)	Resistify v1.1.1; SignalP 6.0	NLR and PRR	592 NLR; 838 PRR; 134 canonical CNL	222 resequenced accessions; 36 infected-fruit libraries
Ngou et al., 2024 [4]	Land plants	Phylogenomics of PRR families	PRR	Family-level	No population analysis
Feng et al., 2025 [6]	Capsicum annuum	Genome-wide NLR identification and evolution	NLR	Family-level	Pathogen-responsive expression
Chen et al., 2026 [5]	Morus spp.	Pan-genome immune-receptor characterization	NLR, RLK and RLP	Family-level	Pan-genome; pathogen stress
Steuernagel et al., 2020 [7]	Angiosperms	NLR-Annotator (sequence-based)	NLR	Method	Not applicable

4.2. Contrasting Variation Between the Two Families

Nucleotide diversity was higher in PRR than in NLR genes in both the filtered and the full window-based analyses. This contrast should be interpreted in the light of the different functional constraints acting on the two families. Because PRRs recognize conserved microbial patterns and act as the first line of surveillance, many of their genes may retain standing variation; NLR genes, by contrast, are engaged in a direct molecular arms race, and their maintenance is often achieved through presence/absence polymorphism and copy-number variation rather than through high nucleotide diversity at individual loci. The magnitude of the difference, however, depended on the filtering scheme (5.1-fold in the MAF-filtered subset versus 1.47-fold in the full window analysis), so we do not interpret the absolute values as a quantitative measure of diversity. The promoter analysis is consistent with a difference in regulatory architecture, but W-box abundance did not predict response magnitude, and the two families should not be treated as independent layers: cell-surface and intracellular receptors share signaling components and can mutually potentiate immunity [11–15].

4.3. No Pronounced Geographic Structure at Immune Receptor Loci

PCA, phylogenetic trees and split networks did not resolve the 222 accessions into geographically defined groups, and pairwise *F*ST among the 10 populations with $n \geq 3$ was ≤ 0 for all 45 comparisons. This is consistent with the weak genome-wide structure reported for the same germplasm [10], but our data do not identify its demographic cause. The reticulate topology of the split networks is compatible with recombination, incomplete lineage sorting or limited resolution; it is not, by itself, evidence of historical gene flow. We therefore do not infer that resistance alleles are broadly distributed across regions. If the absence of structure is confirmed in a joint re-genotyped dataset, global sampling of germplasm for resistance screening would be justified; the present data support only a descriptive statement of no detectable geographic structure at these loci.

4.4. The PRR Layer Is More Transcriptionally Responsive than the NLR Layer During Infection

The most consistent functional result of this study is the asymmetry between the two families during *C. gloeosporioides* infection. Relative to the 19,548 testable genes, NLR genes were under-represented among differentially expressed genes, whereas PRRs were significantly enriched at 2 dpi; 45 of the 51 genes that responded in at least two contrasts were PRRs, and PRR transcripts generally increased as fungal load rose. This is not merely a power difference: both families were represented by comparable absolute numbers of testable genes (219 NLR and 362 PRR genes), and the NLR set showed a systematic, statistically significant depletion rather than a non-significant result. The conclusion nevertheless applies only to the detectable fraction of each family in fruit tissue, and it concerns transcript abundance, not receptor activity.

Two non-exclusive explanations can be envisaged. First, effective NLR-mediated immunity may be established at the protein level, through effector recognition, oligomerization and resistosome formation, and may even be accompanied by negative feedback on NLR transcription, so that a transcriptional increase is neither necessary nor expected [41]. Second, the defense strategies of red and green fruit may differ qualitatively: [16] attributed the higher resistance of red mango fruit to pre-formed antifungal compounds such as flavonols, anthocyanins and mangiferin rather than to induced transcriptional reprogramming. Our interaction analysis is consistent with this interpretation: only 38 genes (2 dpi) and 333 genes (7 dpi) showed a fruit-type-dependent response, and, among the 7 dpi interactors, NLR and PRR genes were represented at 0.60% and 0.37%, both below the genome-wide background of

1.92%. We did not detect strong evidence for differential induction of most detectable receptor genes between red and green fruit; this is not proof that such differences are absent.

The fungal data provide a complementary view. In the most resistant combination (red fruit, red side), the fungus accumulated only ~1.4% of reads at 7 dpi and expressed a much smaller small-secreted-protein program (24 candidates induced ≥ four-fold) than in susceptible combinations (56–130 candidates). The correlation between host defense gene expression and fungal load further indicates a dose-dependent host response rather than a categorical resistant/susceptible transcriptional state. We therefore present a working model in which the outcome of *C. gloeosporioides* infection in mango fruit is determined jointly by pre-formed chemical defenses and by the inducibility of the PRR layer, with the NLR layer acting largely downstream and post-transcriptionally. The mechanisms of PRR–NLR cooperation and NLR activation have been demonstrated in other systems [11,14,15,41]; our data are consistent with this framework but do not demonstrate it in mango.

4.5. Limitations

Four limitations must be stated explicitly.

First, the population-genetic analyses rest on a merged variant set with a very low genotype call rate. Independent quality control revealed mean per-site call rates of only 4.04% (NLR) and 8.34% (PRR) in the complete extracted sets; after strict filtering (call rate ≥ 80%, biallelic, MAF ≥ 5%), 742 NLR and 2149 PRR sites remained, with mean call rates of 91.9% and 92.4%. The original per-sample VCF and BAM files were produced on a server that is no longer accessible, and the 222 resequencing libraries (approximately 625 GB) were not re-genotyped within this revision. The reduction from 147,199 NLR and 101,605 PRR extracted SNPs to 742 and 2149 retained sites is substantial; consequently, nucleotide diversity and population structure are interpreted qualitatively, and the absolute values and fine-scale topology should not be over-interpreted. The filtered analysis is therefore a conservative reanalysis of the available genotypes, not a replacement for joint re-genotyping. The absolute π values are sensitive to MAF filtering and should not be compared with unfiltered genome-wide estimates; the family contrast, the absence of detectable structure and the negative *F*ST values are the robust observations. Joint re-genotyping (for example, GLnexus or bcftools merge −0 followed by uniform filtering) remains the preferred route for a definitive population analysis.

Second, several population groups are represented by very few accessions (Israel, $n = 2$; Brazil, Mexico and Sri Lanka, $n = 1$). These groups were excluded from pairwise *F*ST and from the statistical comparison of π; their values are reported only descriptively in Table S8. In addition, the 838 PRR genes are computational candidates. A signal peptide does not establish cell-surface localization, and not every RLK or RLP functions as a pathogen-recognition receptor; the PRR count should therefore be read as a candidate repertoire rather than as a validated receptor inventory.

Third, the transcriptional conclusions apply to the fraction of the two families that is expressed in mango fruit. Only 219 of 592 NLR genes (37.0%) and 362 of 838 PRR genes (43.2%) passed the expression filter, because the libraries are single-end, 61 bp and of modest depth (median 2.9 million reads). Fungal load was inferred from the proportion of reads assigned to the fungal genome and was not validated here by quantitative PCR, lesion measurements or microscopy. The *C. gloeosporioides* isolate used for the secreted-protein predictions (Nara gc5) is not the isolate used in the inoculation experiment, so the expression estimates rely on mapping against a heterologous reference and should be treated as approximate. Finally, we did not run EffectorP 3.0 or structural effector comparisons, and no functional validation of NLR, PRR or fungal candidates was performed. The 1035 small secreted proteins (<400 aa) are prioritized candidates for future work, not demonstrated effectors.

4.6. Implications for Breeding

Three practical implications follow, all of which require confirmation. First, the CNL-dominated and strongly clustered NLR repertoire of mango, together with the particularly NLR-rich chromosome 2 and Scaffold2 regions, provides a defined target set for association mapping and for the search for presence/absence variation underlying anthracnose resistance. Second, the observation that PRR genes carry the more diverse and more transcriptionally responsive component of the detectable immune-receptor repertoire, and that PRR promoters are enriched in W-box elements, singles out the PRR layer and WRKY-associated regulation as promising entry points for functional studies. Third, if the absence of geographic structure at immune-receptor loci is confirmed in a joint re-genotyped dataset, resistance pre-breeding could sample germplasm globally rather than regionally; the present data do not by themselves justify that recommendation.

4.7. Conclusions

We report a genome-wide inventory of NLR and PRR immune receptor genes in mango: 592 NLR genes (1879 transcripts) and 838 PRR genes (1049 transcripts), including 134 genes with a canonical CNL architecture. Using a strictly filtered genotype matrix (742 NLR and 2149 PRR sites), we found higher nucleotide diversity in PRR than in NLR genes, no detectable geographic structure, and pairwise *F*ST estimates ≤ 0 among populations with $n \geq 3$. In infected fruit, 45 of the 51 genes responsive in at least two contrasts were PRRs, and NLR genes were under-represented among differentially expressed genes; PRR promoters were enriched in W-box elements, but W-box abundance did not predict response magnitude. Under the conditions studied, detectable PRR genes were more transcriptionally responsive than NLR genes, suggesting possible differences in regulation rather than a demonstrated functional hierarchy. These results provide a genome-wide resource and testable hypotheses for mango immunity, and they define the re-genotyping and functional experiments needed to turn the population and expression patterns into breeding targets.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes17121524/s1. Supplementary figures: Figure S1: variant-calling quality control; Figure S2: RNA-seq quantification quality control; Figure S3: sample-to-sample distance; Figure S4: expression of core responsive genes; Figure S5: fruit type by infection interaction; Figure S6: SNP phylogeny and split networks; Figure S7: population differentiation; Figure S8. Full-window nucleotide diversity (π) of NLR and PRR genes, shown as a sensitivity analysis for Figure 4. Supplementary tables: Table S1: accession information; Table S2: population composition; Table S3: chromosomal distribution; Table S4: NLR inventory; Table S5: PRR inventory; Table S6: variant annotation; Table S7: genotype-calling quality; Table S8: nucleotide diversity; Table S9: pairwise differentiation; Table S10: RNA-seq library summary; Table S11: differential expression summary; Table S12: family enrichment; Table S13: core responsive genes; Table S14: fungal load; Table S15: promoter cis-elements; Table S16: secreted-protein candidates; Table S17: W-box copy number versus transcriptional response magnitude.

Author Contributions: B.J., J.D. and M.B. contributed equally to this work. Conceptualization, J.D., Y.Z. and X.Z.; Methodology, B.J., J.D. and M.B.; Software, B.J.; Validation, M.B., B.S., M.G. and K.S.; Formal Analysis, B.J., J.D. and M.B.; Investigation, B.S., M.G. and K.S.; Resources, Y.Z. and X.Z.; Data Curation, B.J., J.D. and K.S.; Writing—Original Draft Preparation, B.J., J.D. and M.B.; Writing—Review & Editing, B.J., J.D., M.B., B.S., M.G., K.S., Y.Z. and X.Z.; Visualization, B.J. and M.B.; Supervision, Y.Z. and X.Z.; Project Administration, Y.Z. and X.Z.; Funding Acquisition, Y.Z. and X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The resequencing data are available in the Genome Sequence Archive under accession CRA016056 (BioProject PRJCA025449; https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA025449, accessed on 9 October 2026). The infection transcriptome is available in the NCBI Sequence Read Archive under BioProject PRJNA575336 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA575336, accessed on 9 October 2026). The *C. gloeosporioides* Nara gc5 genome and annotation are available under GCA_000319635.1 (https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_000319635.1/, accessed on 9 October 2026). Source data for all main and supplementary figure panels are provided in the Supplementary Materials. All other data supporting the conclusions are included in the article and its Supplementary Materials. The authors thank the public data providers: the Genome Sequence Archive (CRA016056), NCBI SRA (PRJNA575336), and NCBI RefSeq (GCA_000319635.1). Computational resources were provided by the High-Performance Computing platform of Yunnan Agricultural University.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zipfel, C. Plant pattern-recognition receptors. *Trends Immunol.* **2014**, *35*, 345–351. [CrossRef] [PubMed]
2. Jones, J.D.G.; Vance, R.E.; Dangl, J.L. Intracellular innate immune surveillance devices in plants and animals. *Science* **2016**, *354*, aaf6395. [CrossRef] [PubMed]
3. Tahir ul Qamar, M.; Sadaqat, M.; Zhu, X.-T.; Li, H.; Huang, X.; Fatima, K.; Almutairi, M.M.; Chen, L.-L. Comparative genomics profiling revealed multi-stress responsive roles of the CC-NBS-LRR genes in three mango cultivars. *Front. Plant Sci.* **2023**, *14*, 1285547. [CrossRef] [PubMed]
4. Ngou, B.P.M.; Wyler, M.; Schmid, M.W.; Kadota, Y.; Shirasu, K. Evolutionary trajectory of pattern recognition receptors in plants. *Nat. Commun.* **2024**, *15*, 308. [CrossRef] [PubMed]
5. Chen, Y.; Fan, W.; Xu, M.; Peng, H.; Li, Q.; Liu, D.; Xia, Z.; Wu, Z.; Wang, X.; Zhao, A. Pan-genomic stratification and multi-dimensional regulatory characterization of immune receptors suggest growth-defense co-regulation during development and pathogen stress in mulberry (*Morus* spp.). *Plant Stress* **2026**, *24*, 101550. [CrossRef]
6. Feng, C.; Chen, Q.; Liu, W.; Li, T.; Ji, T. Genome-wide identification and functional evolution of NLR gene family in Capsicum annuum. *Curr. Issues Mol. Biol.* **2025**, *47*, 867. [CrossRef] [PubMed]
7. Steuernagel, B.; Witek, K.; Krattinger, S.G.; Ramirez-Gonzalez, R.H.; Schoonbeek, H.-J.; Yu, G.; Baggs, E.; Witek, A.I.; Yadav, I.; Krasileva, K.V.; et al. The NLR-Annotator tool enables annotation of the intracellular immune receptor repertoire. *Plant Physiol.* **2020**, *183*, 468–482. [CrossRef] [PubMed]
8. Zeng, Z.; Li, S.-X.; Wu, W.-S.; Zhao, P.; Shao, Z.-Q.; Liu, Y. The landscape of integrated domains of angiosperm NLR genes reveals continuous architecture evolution of plant intracellular immune receptors. *Plants* **2026**, *15*, 81. [CrossRef] [PubMed]
9. Wang, P.; Luo, Y.; Huang, J.; Gao, S.; Zhu, G.; Dang, Z.; Gai, J.; Yang, M.; Zhu, M.; Zhang, H.; et al. The genome evolution and domestication of tropical fruit mango. *Genome Biol.* **2020**, *21*, 60. [CrossRef] [PubMed]
10. Ma, X.; Wu, H.; Liu, B.; Wang, S.; Zhang, Y.; Su, M.; Zheng, B.; Pan, H.; Du, B.; Wang, J.; et al. Genomic diversity, population structure, and genome-wide association reveal genetic differentiation and trait improvements in mango. *Hortic. Res.* **2024**, *11*, uhae153. [CrossRef] [PubMed]
11. Ngou, B.P.M.; Ding, P.; Jones, J.D.G. Thirty years of resistance: Zig-zag through the plant immune system. *Plant Cell* **2022**, *34*, 1447–1478. [CrossRef] [PubMed]
12. Sperschneider, J.; Dodds, P.N. EffectorP 3.0: Prediction of apoplastic and cytoplasmic effectors in fungi and oomycetes. *Mol. Plant-Microbe Interact.* **2022**, *35*, 146–156. [CrossRef] [PubMed]
13. Seong, K.; Krasileva, K.V. Prediction of effector protein structures from fungal phytopathogens enables evolutionary analyses. *Nat. Microbiol.* **2023**, *8*, 174–187. [CrossRef] [PubMed]
14. Yuan, M.; Jiang, Z.; Bi, G.; Nomura, K.; Liu, M.; Wang, Y.; Cai, B.; Zhou, J.-M.; He, S.Y.; Xin, X.-F. Pattern-recognition receptors are required for NLR-mediated plant immunity. *Nature* **2021**, *592*, 105–109. [CrossRef] [PubMed]
15. Feehan, J.M.; Wang, J.; Sun, X.; Choi, J.; Ahn, H.-K.; Ngou, B.P.M.; Parker, J.E.; Jones, J.D.G. Oligomerization of a plant helper NLR requires cell-surface and intracellular immune receptor activation. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2210406120. [CrossRef] [PubMed]
16. Sudheeran, P.K.; Sela, N.; Carmeli-Weissberg, M.; Ovadia, R.; Panda, S.; Feygenberg, O.; Maurer, D.; Oren-Shamir, M.; Aharoni, A.; Alkan, N. Induced defense response in red mango fruit against *Colletotrichum* gloeosporioides. *Hortic. Res.* **2021**, *8*, 17. [CrossRef] [PubMed]

17. Smith, M.; Jones, J.T.; Hein, I. Resistify: A Novel NLR Classifier That Reveals Helitron-Associated NLR Expansion in Solanaceae. *Bioinform. Biol. Insights* **2025**, *19*, 11779322241308944. [CrossRef] [PubMed]
18. Teufel, F.; Almagro Armenteros, J.J.; Johansen, A.R.; Gíslason, M.H.; Pihl, S.I.; Tsirigos, K.D.; Winther, O.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 6.0 predicts all five types of signal peptides using protein language models. *Nat. Biotechnol.* **2022**, *40*, 1023–1025. [CrossRef] [PubMed]
19. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
20. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately Maximum-Likelihood Trees for Large Alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef] [PubMed]
21. Chen, S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta* **2023**, *2*, e107. [CrossRef] [PubMed]
22. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997. [CrossRef]
23. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.A. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
24. Poplin, R.; Chang, P.-C.; Alexander, D.; Schwartz, S.; Colthurst, T.; Ku, A.; Newburger, D.; Dijamco, J.; Nguyen, N.; Afshar, P.T.; et al. A universal SNP and small-indel variant caller using deep neural networks. *Nat. Biotechnol.* **2018**, *36*, 983–987. [CrossRef] [PubMed]
25. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCFtools. *GigaScience* **2021**, *10*, giab008. [CrossRef] [PubMed]
26. Wang, K.; Li, M.; Hakonarson, H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **2010**, *38*, e164. [CrossRef] [PubMed]
27. Ortiz, E.M. vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for phylogenetic analysis. *Zenodo*. 2019. Available online: https://zenodo.org/records/2540861 (accessed on 9 October 2026).
28. Minh, B.Q.; Schmidt, H.A.; Chernomor, O.; Schrempf, D.; Woodhams, M.D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37*, 1530–1534. [CrossRef] [PubMed]
29. Zhou, T.; Xu, K.; Zhao, F.; Liu, W.; Li, L.; Hua, Z.; Zhou, X. itol.toolkit accelerates working with iTOL (Interactive Tree of Life) by an automated generation of annotation files. *Bioinformatics* **2023**, *39*, btad339. [CrossRef] [PubMed]
30. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef] [PubMed]
31. Huson, D.H.; Bryant, D. Application of Phylogenetic Networks in Evolutionary Studies. *Mol. Biol. Evol.* **2006**, *23*, 254–267. [CrossRef] [PubMed]
32. Danecek, P.; Auton, A.; Abecasis, G.; Albers, C.A.; Banks, E.; DePristo, M.A.; Handsaker, R.E.; Lunter, G.; Marth, G.T.; Sherry, S.T.; et al. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]
33. Pertea, G.; Pertea, M. GFF Utilities: GffRead and GffCompare. *F1000Research* **2020**, *9*, 304. [CrossRef] [PubMed]
34. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **2017**, *14*, 417–419. [CrossRef] [PubMed]
35. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [CrossRef] [PubMed]
36. Li, H. Minibwa: A Faster And More Accurate Successor of BWA-MEM. GitHub Repository, 2024. Available online: https://github.com/lh3/minibwa (accessed on 9 October 2026).
37. Langmead, B.; Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359. [CrossRef] [PubMed]
38. Thumuluri, V.; Almagro Armenteros, J.J.; Johansen, A.R.; Nielsen, H.; Winther, O. DeepLoc 2.0: Multi-label subcellular localization prediction using protein language models. *Nucleic Acids Res.* **2022**, *50*, W228–W234. [CrossRef] [PubMed]
39. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. [CrossRef] [PubMed]
40. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [CrossRef] [PubMed]
41. Maruta, N.; Sorbello, M.; Garzon-Flores, L.; Kobe, B. Molecular mechanisms of plant NLR activation and signalling. *Plant J.* **2026**, *125*, e70702. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.