

Integrative approach to identify robust Pattern Recognition Receptors in *Eucalyptus grandis*: novel candidates for disease resistance

Yesica Elisa Gonzalez

gonzalez.yesica@inta.gob.ar

Instituto de Agrobiotecnología y Biología Molecular

Natalia Cristina Aguirre

Instituto de Agrobiotecnología y Biología Molecular

Carla Valeria Filippi

Universidad de la República

Research Article

Keywords:

Posted Date: December 29th, 2025

DOI: <https://doi.org/10.21203/rs.3.rs-8407889/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Plant Molecular Biology on June 11th, 2026. See the published version at <https://doi.org/10.1007/s11103-026-01734-6>.

Abstract

Pattern-recognition receptors (PRRs) initiate plant pattern-triggered immunity (PTI), encompassing receptor-like kinases (RLKs) and receptor-like proteins (RLPs). *Eucalyptus grandis*, an economically important species worldwide, is a long-lived organism that faces multiple disease pressures. The deployment of PRR-based breeding tools offers a route to broad, non-race-specific resistance that can remain effective across outbreaks. The objective of this work was to identify and characterise the PRR repertoire of *E. grandis* using public multi-omics data. Two state-of-the-art predictors were compared with a multi-species curation and a meta-analysis was conducted compiling transcriptomic data under biotic stress. A set of 730 PRR candidates were consistently identified across three sources (~ 1.6% of protein-coding genes), of which 283 were differentially expressed (PRR-DEGs). A PFAM domain-based scheme was applied to standardise classification across tools. RLKs (TM-K-LRR) predominated over RLPs (TM-LRR), while many carried leucine-rich repeats or leucine-rich repeat-like ectodomains. PRR genes were unevenly distributed across the genome: chromosome 6 had the highest count and the densest clustering. Gene-family expansion appeared mainly driven by duplication, with extensive tandem arrays supported by segmental, proximal and dispersed events. Gene Ontology and cis-elements annotations in PRR/PRR-DEGs showed significant enrichment in terms related to cell cycle development, hormonal regulation and stress response. We proposed a catalogue of 16 PRR which resulted in DEGs against at least three pathogens, suggesting their broad spectrum and robustness. Most of them presented orthologues with cross-taxon evidence of defence roles. This study delineates novel multi-pathogen candidate PRR genes, providing valuable information to assist *Eucalyptus* breeding programs.

1. Introduction

The *Eucalyptus* genus, which belongs to the Myrtaceae family, is native to Australia, New Guinea, Timor, Indonesia and the Philippines, and comprises over 700 species and subspecies (Ladiges et al., 2003; Thornhill et al., 2019). Due to traits such as rapid growth, short rotation cycles, year-round harvesting, and high adaptability and productivity (Gullón et al., 2020), *Eucalyptus* has become widely cultivated across tropical, subtropical, and temperate regions worldwide, covering over 20 million hectares. It plays a key role in the global hardwood forestry industry, with applications in timber, pulp and paper and biofuels.

Among the cultivated species, *Eucalyptus grandis* stands out for its fast growth and early yield, making it particularly valuable in commercial plantations (Naidoo et al., 2014). However, despite its economic importance, *E. grandis* is highly susceptible to several biotic and abiotic stresses that compromise growth and wood quality. Biotic threats include myrtle rust caused by *Austropuccinia psidii* (syn. *Puccinia psidii*); the stem canker pathogen *Chrysosporthe austroafricana*; root rot pathogen *Phytophthora cinnamomi*; leaf blight caused by *Calonectria* spp.; and pests such as *Leptocybe invasa* (Naidoo et al., 2014). These threats have intensified in recent decades due to climate change, which promotes disease emergence by extending infection windows, increasing pathogen virulence, and expanding host ranges (Younessi-Hamzekhanlu & Gailing, 2022).

Current management strategies include the use of tolerant *Eucalyptus* genotypes and integrated pest management, such as biological control for *L. invasa* (Naidoo et al., 2014). However, these measures often fall short, underscoring the need for improved resistance strategies through breeding and biotechnological approaches (Wingfield et al., 2013). In this context, understanding the molecular mechanisms of plant immunity is essential for developing more resilient varieties.

A major breakthrough for *Eucalyptus* research was the sequencing of the *E. grandis* genome, with > 94% of the genome anchored to chromosomes, which has enabled omics-based studies and large-scale analysis of gene families involved in defence responses (Myburg et al., 2014; Zwart et al., 2017). This genomic resource has established *E. grandis* as a genomic model for studying the defence mechanisms of woody perennials, benefiting not only this species but also other members of the Myrtaceae family (Christie et al., 2015).

In plants, the first line of inducible immune defence, known as pattern-triggered immunity (PTI), is initiated when pattern recognition receptors (PRRs) located on the plasma membrane detect pathogen-, microbe-, or damage-associated molecular patterns (PAMPs, MAMPs, or DAMPs) (Jones & Dangl, 2006; Yuan et al., 2021). Two main classes of PRRs are recognised: receptor-like kinases (RLKs) and receptor-like proteins (RLPs), which play a pivotal role in sensing pathogens and maintaining cellular homeostasis (Silva et al., 2022).

RLKs are characterised by a typical tripartite structure: an extracellular domain (ectodomain) at the amino-terminal (N-terminal) region, a transmembrane domain (TM), and an intracellular serine/threonine kinase domain at the carboxy-terminal (C-terminal) region. The ectodomain, responsible for ligand perception, can include leucine-rich repeats (LRR), the lysine motif (LysM), the lectin domain (Lec), and/or an epidermal growth factor-like domain (EGF) (Liu et al., 2023). This ectodomain interacts with PAMPs/MAMPs or DAMPs, while the kinase domain activates intracellular signalling through phosphorylation (Silva et al., 2022). RLKs are grouped into families based on ectodomain structure, with LRR-RLKs being the most abundant, followed by lectin-type RLKs (Minkoff et al., 2017).

RLPs, in contrast, share similar ectodomains with RLKs but lack the intracellular kinase domain. Their signalling relies on forming complexes with RLKs or with receptor-like cytoplasmic kinases (RLCKs), which transduce the signal into the cell (Liu et al., 2023). The functional characterisation of RLPs has proven more challenging, as their variable domains and lack of conserved kinase motifs hinder identification through standard homology-based methods (Silva et al., 2022).

Given the importance of PRRs in the early stages of immune response, their identification is a crucial step in the genetic improvement of disease-resistant plants. Although experimental approaches can confirm PRR function, they are time-consuming and resource-intensive. In this regard, bioinformatics tools for *in silico* prediction and classification of PRRs have become increasingly valuable to accelerate the discovery and annotation process (Del Hierro et al., 2021).

This study addresses the identification and characterisation of PRR genes involved in PTI in *E. grandis* by integrating computational predictions with functional analysis. Two state-of-the-art prediction tools: RRGPredictor (Santana Silva & Micheli, 2020) and DRAGO3 (Calle García et al., 2022) were employed, and their outputs were compared with a recent large-scale PRR dataset derived from 350 plant species (Ngou et al., 2022). In addition, we conducted a multi-layered characterisation that included differential expression meta-analysis, updated functional annotation, phylogenetic and evolutionary analysis, and promoter region examination to identify cis-regulatory elements (CREs). This integrative approach aims to shed light on the diversity, structure, and potential regulatory mechanisms of PRRs in *E. grandis*, contributing to a better understanding of defence responses in woody species.

2. Materials and Methods

2.1 Genomic, proteomic, and transcriptomic resources

This study used publicly available omics data for *E. grandis*, retrieved from the Phytozome database¹ (Goodstein et al., 2012). The resources included genome, transcriptome, and proteome data corresponding to version 2.0 of the *E. grandis* reference genome (Myburg et al., 2014; Bartholomé et al., 2015), which incorporates structural and functional annotations.

2.2 Identification of PRR genes

To identify PRR genes involved in plant defence, we implemented three complementary bioinformatics strategies: 1) DRAGO3 (Disease Resistance Analysis and Gene Orthology version 3; Calle García et al., 2022) was used to scan the *E. grandis* v2.0 proteome; 2) RRGPredictor (Santana Silva & Micheli, 2020) was applied after analysing the proteome with InterProScan5 (Jones et al., 2014); and 3) filtering the *E. grandis* only PRR gene set reported by Ngou et al. (2022).

As DRAGO3 and RRGPredictor permit the annotation of resistance genes in general, we filtered only PRRs by retaining proteins with transmembrane (TM) domains predicted by TMHMM 2.0 (Krogh et al., 2001) and without nucleotide binding site (NBS) domains, which are specific of R genes (Christie et al., 2015). Redundancy among predicted sequences was reduced using CD-HIT v4.7 (Li & Godzik, 2006), with a 90% sequence identity threshold, retaining only the longest protein.

Genes consistently predicted across the three strategies were defined as putative PRRs. To harmonise classification across tools, we established a unified domain-based characterisation using PFAM domains identified via InterProScan5, with e-values <0.001 and an alignment coverage greater than 50%. Functionally related domains, such as various leucine-rich repeats (LRRs) and kinase-related motifs, were grouped under "LRR" and "K", respectively.

2.3 Meta-Analysis of RNAseq data

In order to assess the expression dynamics of candidate PRRs, a meta-analysis of RNA-seq datasets derived from *Eucalyptus* species challenged with pathogens was conducted, encompassing studies published between 2014 (*E. grandis* genome release) and March 2023.

Only studies with publicly available raw RNA-seq data were considered, including responses to *C. austroafricana* (Mangwanda et al., 2015), *L. invasa* (Oates et al., 2015), *A. psidii* (Santos et al., 2020), and *Ralstonia solanacearum* (Xiaohui et al., 2022). Datasets were obtained from the National Center for Biotechnology Information repository (NCBI: PRJNA280236²; PRJNA305347³ and PRJNA588626⁴) and the Genome Sequence Archive (PRJCA006666⁵).

Reads were quality-filtered using Trimmomatic v0.36 (Bolger et al., 2014), removing adapters and discarding sequences shorter than 80 bp or with average Phred scores below 30.

Transcript-level quantification was performed with Salmon v0.12.0 (Patro et al., 2017) in quasi-mapping mode, enabling GC bias correction. The mapping index was constructed using the *E. grandis* reference transcriptome (v2.0). Gene-level abundance estimates were summarised using Tximport v1.26.1 (Soneson et al., 2015). Differential gene expression analysis was conducted using DESeq2 v1.38.3 (Love et al., 2014), employing its default pipeline based on a negative binomial regression model.

Exploratory analyses included principal component analysis (PCA) and volcano plots, both of which were visualized using ggplot2 v3.4.2 (Wickham H., 2016).

2.4 Phylogenetic Analysis of PRRs

Putative PRR protein sequences were aligned using ClustalOmega (Sievers & Higgins, 2021). Phylogenetic trees were constructed using the Neighbor-Joining (NJ) method based on pairwise sequence identity matrices. The R packages "msa v1.30.1" (Bodenhofer et al., 2015), "seqinr v4.2-30" (Charif & Lobry, 2007), and "ape v5.7-1" (Paradis & Schliep, 2019) were used for alignment, distance calculation and tree construction, respectively. PRRs were classified into DEGs and non-DEGs based on the RNA-seq meta-analysis results.

2.5 New functional annotation and Orthology

Putative PRR proteins were functionally annotated using PANNZER (Protein ANnotation with Z-score, Törönen & Holm, 2022), and results were compared with the *E. grandis* v2.0 genome annotation. Gene Ontology (GO) enrichment analysis was performed using TopGO v2.50.0 (Alexa & Rahnenfuhrer, 2024),

applying the 'weight01' algorithm to explore overrepresented Biological Process and Molecular Function terms among PRR annotated by both methodologies.

Orthologs in *Arabidopsis thaliana* were identified using PLAZA 5.0 Dictos⁶, enabling comparative insights across species (Van Bel et al., 2022).

2.6 Gene duplication analysis

Gene duplication events were identified with McScanX (Wang et al., 2012), using the amino acid sequences of PRRs. Duplicated gene pairs and their chromosomal locations were visualized using TBtools v.2.119 (Chen et al., 2023).

Non-synonymous (Ka) and synonymous (Ks) substitution rates were estimated for duplicated gene pairs based on coding sequence (CDS) alignments, applying the Nei-Gojobori method (Nei & Gojobori, 1986).

2.7 Analysis of Cis-Regulatory Elements

Sequences 1,500 bp upstream of the start codon (ATG) of putative PRR genes were retrieved to identify cis-regulatory elements (CREs) in the promoter regions, using PlantCare⁷ (Lescot et al., 2002). CREs associated with stress responses were visualised using ggplot2 and their enrichment in DEGs versus non-DEG PRRs was evaluated via Fisher's exact test.

3. Results

3.1 Identification of genes involved in defense mechanisms and PRR

The analysis of the *E. grandis* proteome (46,280 sequences) using DRAGO3 identified 4,976 proteins potentially involved in defence mechanisms. Filtering for proteins containing at least one transmembrane domain (TM) reduced this number to 4,186 proteins. Exclusion of proteins with NBS domains resulted in 3,258 putative PRR proteins. A redundancy reduction step using CD-HIT yielded 2,831 unique putative PRR proteins, corresponding to 2,793 genes. Using a similar workflow, RRGPredictor identified 3,151 putative resistance proteins without NBS motifs. After applying the TM domain filter and CD-HIT redundancy reduction, a total of 1,271 putative PRR proteins were retained, corresponding to 1,261 genes. Additionally, 810 PRR genes and proteins were compiled from the study by Ngou et al. (2022), which defined 19 PRR categories. These included 300 genes as RLK-LRR-XII, 224 as RLP-LRR, 65 as RLK-LRR-XI, 51 as RLK-LRR-VIII, 38 as RLK-LRR-III, 25 each as RLK-LysM and RLK-LRR-I, and smaller numbers in other RLK-LRR subfamilies (Figure 1.A).

Altogether, a total of 2,835 PRR genes were identified by at least one of the three approaches: 2,793 by DRAGO3 (2,800 proteins), 1,261 by RRGPredictor (1,261 proteins), and 810 by Ngou et al. (2022) (810 proteins). DRAGO3 classified these gene products into 19 categories: 1,016 as KIN, 588 as RLK, 581 as RLP, 341 as LECRK, 77 as CK, 42 as TRAN, 39 as T, 35 as LEC, 29 as LYK, 20 as CLK, 7 as LYP, 7 as CL, 5 as LYS, 4 as TL, 4 as C, 2 as CLECRK, and 1 each as CT, CLYSK, and UNDEFINED (Figure 1.A). RRGPredictor assigned these genes to six categories: 802 as UNKNOWN, 269 as RLK, 146 as RLK-GNK2, 30 as MLO, 8 as T, and 6 as RLP (Figure 1.A).

3.1.1 Recategorization of PRR proteins consistently identified by three strategies

A comparison among the three strategies revealed 730 genes identified consistently across all datasets (Supplementary Information 1). These were considered high-confidence PRRs. DRAGO3 classified them into RLK, RLP, CLK, CL, and UNDEFINED; RRGPredictor into UNKNOWN and RLK; and Ngou et al. (2022), into various RLK-LRR and RLP-LRR subgroups. The observed disparity in categorization criteria among the three strategies prompted the development of a unified classification system for PRR proteins and genes. To achieve this, we analysed the presence and absence of protein domains based on PFAM domain annotations from InterProScan5. Only domains passing the filtering criteria (e-value < 0.001 and alignment coverage greater than 50%) were considered. According to the PFAM database (Mistry et al., 2021), the following domains in different combinations were present in the 730 PRRs: TM domains, K endodomains (PFAM: "PF00069" and "PF07714") and among the ectodomains LRR (PFAM: "PF08263", "PF12799", and "PF13855"), Malectin (PFAM: "PF11721"), Malectin-like (PFAM: "PF12819"). Consequently, the 730 putative PRR proteins were grouped into 11 new categories (Figure 1.B).

3.2 Meta-Analysis of RNAseq

A meta-analysis of four published transcriptomic studies identified 12,157 differentially expressed genes (DEGs). Exploratory analyses, including PCA and volcano plots, are shown in Supplementary Information 2, 3, 4, 5. Cross-referencing these DEGs with the 730 putative PRRs revealed 283 genes with evidence of pathogen-responsive expression, hence defined as PRR-DEGs.

Pathogen-specific expression patterns were observed. A total of 70, 147, 106 and 75 PRR-DEGs were detected for *A. psidii* (rust), *R. solanacearum* (bacterial wilt), *C. austroafricana* (canker) and *L. invasa* (gall wasp), respectively. Besides, 28, 90, 47 and 25 were unique for each pathogen, respectively. On the other hand, among the PRR-DEGs, approximately 41% were consistently downregulated and 34% consistently upregulated. The remaining 25% showed variable directionality depending on the pathogen, treatment, or genetic background. Additionally, 18 genes responded to three or all pathogens, suggesting potential broad-spectrum (Figure 2).

3.3 Phylogenetic analysis of PRR and PRR-DEG

To explore relationships between domain composition and functional roles, Neighbor-Joining phylogenetic trees were generated for both the full set of 730 PRRs (Supplementary Information 6) and the subset of 283 PRR-DEGs. Proteins were colour-coded by the new 11 categories, and by their differential expression (or not) against specific pathogens, respectively. An integrated overview tree summarises these results (Figure 3).

Initial analysis revealed clear clustering patterns between RLKs and RLPs. Among the grouped sequences, the TM-LRR category accounted for approximately 25% of the total, notable for its well-defined clustering. Similarly, the various categories within RLKs also showed distinctly defined clusters.

Regarding the pathogen response of PRR-DEGs, it was observed that about 30% of the RLPs were DEGs and only exhibited evidence of response to one or two pathogens. Conversely, approximately 40% of the RLK-related categories were DEGs, showing responses from one to four pathogens. Furthermore, RLKs with malectin domains predominantly responded to more than one pathogen.

3.3.1 Chromosome distribution and clustering of PRR and PRR-DEG

An analysis of gene distribution across chromosomes identified 663 putative PRR genes on the eleven chromosomes and 67 at the scaffold level. Chromosomes 6 and 4 harboured the highest and the lowest number of PRR genes, with 130 and 29 respectively. This finding indicates no direct correlation between the number of PRR genes and chromosome size, as chromosomes 3 and 10 are the longest and the shortest, respectively (Figure 4). Additionally, the proportion of PRR genes per chromosome ranged from 1% on chromosomes 2 and 8 to 3.2% on chromosome 6 (Table 1).

Moreover, PRR genes were classified as within a "Cluster", as genes located less than 200,000 base pairs apart from each other (Holub, 2001), or "Singleton". This analysis revealed 462 genes in clusters and 201 as singletons across the eleven chromosomes. Notably, the majority of clusters were found on chromosome 6, with the minority on chromosome 2.

Among the 283 PRR-DEGs, 265 were identified on the eleven chromosomes and 18 at the scaffold level (Figure 4).

Additionally, chromosomes 10 and 11 exhibited a >50% higher proportion of PRR-DEGs relative to the overall PRR distribution (Table 1).

Tab. 1 Summary of the pattern recognition receptors (PRR) genes identified on the eleven chromosomes and scaffolds. Chromosome: chromosome number; Size in bp: size of the chromosome in base pairs; No. of total genes: number of total genes per chromosome; No. of PRR: number of putative PRR genes identified by the three strategies; % PRR: percentage of PRR is represented based on the total number of coding genes on the chromosome; No. of differentially expressed gene (DEG): the number of PRR-DEG genes per chromosome; No. of Clustered genes: the number of PRR genes that are clustered together (distance < 200 kpb); No. of cluster: the number of clusters per chromosome; No. of Singleton: the number of genes that are not clustered together; No. of Singleton: the number of genes that are not clustered together.

Chromosome	Size in bp	N° of total genes	N° of PRR	% PRR	N° of DEG	N° of Clustered genes	N° of Cluster	N° of Singleton
1	45,527,262	2,698	57	2.1	27	46	9	11
2	60,273,285	3,287	33	1	16	12	5	21
3	85,001,648	3,513	103	3	30	78	12	25
4	41,674,560	2,231	29	1.3	1	18	4	11
5	77,196,103	3,306	63	1.9	21	48	14	15
6	58,190,708	4,054	130	3.2	49	109	23	21
7	55,515,396	2,854	67	2.3	24	44	9	23
8	73,422,429	4,087	41	1	19	23	8	18
9	39,799,183	2,398	45	1.9	15	27	10	18
10	38,249,343	2,663	43	1.6	24	29	9	14
11	45,397,251	3,030	52	1.7	39	28	8	24
Scaffold	-	2,228	67	-	18	-	-	-
Total	640,439,468	36,349	730	2	283	462	111	201

3.4 New functional annotation and GO enrichment

Using the described methodology, GO enrichment graphs for Biological Processes and Molecular Function were generated utilising both the original annotation (*E. grandis* v2.0 reference genome) and the new annotation (with PANNZER tool) of PRR genes. It was generally observed that the new annotation revealed a greater number of significant GO enriched terms that were related to various biotic stresses.

For "Biological Processes", the original annotation of PRRs only yielded the significantly enriched term "protein phosphorylation". Although the phosphorylation process corresponds to the kinase domains of the RLKs, no other GO terms related to the plant Inducible Immunity System were assigned. In contrast, the new annotation identified not only the terms "phosphorylation" (go id: 0016310) and "protein phosphorylation" (go id: 0006468) but also significant terms specific to responses to various biotic stresses, such as "defence response" (go id: 0006952), "defence/response to other organism" (go id: 0098542), and terms specifically related to fungal response like "defence/response to fungus" (go id: 0050832). Additionally, terms related to PTI and PRR responses, such as "transmembrane receptor protein serine/threonine kinase" (go id: 0007178), were identified. (Figure 5.A)

When all putative PRRs and PRR-DEGs were compared based on the new annotation, terms such as "defence response", "response to other organisms" and "response to fungus" remained significant and abundant among the PRR-DEGs. Notably, the term "positive regulation of defence response" also emerged as significant (Figure 5.B).

With regard to the orthologs identified in *A. thaliana*, their functional annotation was found to be consistent with defence responses to a range of stresses.

Regarding "Molecular Function", although the most significantly enriched and abundant terms corresponded to general molecular functions, other significant terms were also observed that could be specifically associated with functions involved in PTI. These included receptor-related functions such as "signaling receptor activity" (go id: 0038023), "receptor serine/threonine kinase binding" (go id: 0033612), "protein serine/threonine kinase activity" (go id: 0004674), and "transmembrane signaling receptor activity" (go id: 0004888) (Supplementary Information 7).

3.5 Analysis of Cis-Regulatory Elements

The analysis of promoter regions located 1500 bp upstream of the coding region was conducted to examine cis-regulatory elements (CREs) in PRR genes. Elements related to cell cycle development, hormonal regulation, and stress response were identified (Figure 6.A). Among these, some genes exhibited cis-elements associated with hormonal responses, such as to methyl jasmonate (MeJA), a derivative of jasmonic acid, abscisic acid (ABA), and other phytohormones like gibberellin, auxin, and salicylic acid (SA). Other stress-related elements included regulatory elements responsive to light, drought, low temperatures, anaerobic conditions, defence, and pathogen response.

Furthermore, the enrichment analysis of CREs revealed that, of a total of 26 CREs significantly enriched (p -value < 0.05) in PRR-DEGs compared to those without DEG evidence to date, the most significantly enriched were elements responsive to auxin and drought (Supplementary Information 8).

To delve into potential transcriptional regulatory mechanisms of PRRs, an additional analysis was conducted focused on PRR-DEGs showing responses to 3 and 4 pathogens. The results showed enriched cis-elements linked to stress responses and hormonal regulation. Each gene exhibited a unique configuration of these elements, with a predominance of those responsive to light and MeJA (Figure 6.B). However, less frequent but notable elements were also identified, such as those associated with pathogen defence, drought and low temperature stress responses, as well as elements associated with anaerobic and anoxic stress responses.

3.6 Gene duplication

Using the MCScanX software, PRR classification based on duplication mechanisms revealed 662 genes categorised into four groups: 1) 122 segmental duplicates or related to whole-genome duplication (WGD), where paralogs are found in collinear blocks within the same species, relating to more ancestral duplications; 2) 128 dispersed duplicates, characterised by a lack of proximity between paralogs on chromosomes or conservation of synteny; 3) 318 tandem duplicates are noted for their location side by side and relation to more recent duplication events in genome evolution; 4) 94 proximal duplicates, where paralogs are close together but separated by fewer than 20 genes.

Based on the segmental duplicates identified by MCScanX (74 pairs), the nonsynonymous to synonymous substitution ratio (K_a/K_s) was calculated. For 47 gene pairs, the K_a/K_s ratio was less than 1 (ranging from 0.051 to 0.597), indicating that these PRRs have undergone purifying (negative) selection. In addition, 27 of these pairs showed a synonymous substitution ratio (pS) greater than 0.75, suggesting that the sequence divergence was sufficient to estimate a reliable synonymous substitution rate (kS).

Additionally, an analysis of duplicate pairs was performed, focusing on genes with evidence as DEGs and those without. Among these gene pairs, 24 lacked evidence of response to biotic stress. In 14 pairs, both genes were DEGs, while in 36 pairs, at least one gene was identified as DEG (Figure 7).

3.7 Catalogue of PRR genes of potential interest in *Eucalyptus* breeding programs

Based on our results, we propose a catalogue of 16 PRR genes that were differentially expressed in at least three diseases, identified in multiple publications, and upregulated in response to at least one pathogen (Table 2). Among these, four genes responded to all four pathogen types analyzed—bacteria, two distinct fungal pathogens, and the gall wasp. Additionally, nine PRR-DEGs, while not responsive to all pathogens included in the meta-analysis, were induced by a range of organisms, including fungi, insects, and bacteria. The remaining three genes responded to two types of fungi as well as bacteria, highlighting the potential broad-spectrum defense capability of these PRRs against diverse biotic stresses.

Regarding the chromosomal distribution of the 16 PRR-DEGs included in the catalogue, two were annotated in scaffolds, while the remaining 14 were distributed across seven of the eleven chromosomes (2, 3, 6, 8, 9, 10, and 11). Notably, twelve potentially broad-spectrum PRR-DEGs were grouped into nine clusters located on chromosomes 3, 6, 8, 10, and 11. Four of these clusters consisted exclusively of PRRs belonging to the same category. On chromosome

6, one cluster comprised RLKs with malectin domains; similarly, chromosome 8 contained another RLK cluster. In addition, two clusters located on chromosome 11 were composed of RLK categories.

Consistent with these findings, all candidate genes had orthologues in *Arabidopsis thaliana*, which themselves have been reported to respond to multiple biotic stresses. Moreover, these orthologous groups are conserved and stress-responsive across a variety of plant species, including forest trees such as pine (*Pinus pinaster*) and olive (*Olea europaea*), as well as citrus species such as lemon (*Citrus limon*) and grapefruit (*Citrus maxima*) (Table 2).

Tab. 2 Catalogue of PRR genes of potential interest in *Eucalyptus* breeding programs. The table summarises 16 PRR genes of interest, including: gene identifier (Gene_ID_V2), chromosome (CHR), genomic position (Start, End), cluster assignment within 200 kbp (N° Cluster (<200 kbp)), pathogens for which the gene shows differential expression (Pathogen), condition of differential expression (Condition DEG), unified PRR category (Categories), orthologues in *A. thaliana* (*A. thaliana* Orthologues), functional annotation in *E. grandis* v2.0 (Functional Annotation (*E. grandis*)), updated functional annotation (New annotation), and supporting literature (Evidence).

Gene_ID_V2	CHR	Start	End	N° Cluster (<200 kpb)	Pathogen	Condition DEG	Categories	Annotation in <i>E. grandis</i> v 2.	
								A. thaliana Orthologous	Fu An
Eucgr.B03087	02	51188519	51192057	Singleton	<i>C. austroafricana</i>	DOWN (R,S)	TM_K_LRR	AT3G25560	NS kin
					<i>L. invasa</i>	DOWN (R)			
					<i>R. solanacearum</i>	DOWN/UP			
Eucgr.C00984	03	15768031	15771767	16	<i>C. austroafricana</i>	UP (R,S)	TM_K_LRR	AT4G29990	Pro far wil ric do
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	UP			
Eucgr.F00314	06	5586470	5590029	47	<i>L. invasa</i>	DOWN (R)	TM_K_LRR	AT4G08850	Le rep like kin pro
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.F00316	06	5561785	5571145	47	<i>L. invasa</i>	DOWN (S)	TM_K_LRR	AT4G08850	Le rep like kin pro
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.F01307	06	17466429	17483294	52	<i>C. austroafricana</i>	DOWN (R)	TM_K_LRR_Malec	AT3G14840	Le rep tra pro
					<i>L. invasa</i>	DOWN (S)			
					<i>R. solanacearum</i>	DOWN/UP			
Eucgr.F04434	06	56878674	56890048	67	<i>C. austroafricana</i>	UP (R)	TM_K_Malec	AT1G56130	Le rep tra pro
					<i>L. invasa</i>	DOWN (R,S)			
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	UP			
Eucgr.H00707	08	9621912	9625470	77	<i>C. austroafricana</i>	DOWN (R,S)	TM_K_LRR	AT4G08850	Le rep like kin pro
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.H00710	08	9652678	9656274	77	<i>C. austroafricana</i>	DOWN (R,S)	TM_K_LRR	AT4G08850	Le rep like kin pro
					<i>L. invasa</i>	DOWN (R)			
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.H00949	06	7852826	7856009	49	<i>C. austroafricana</i>	DOWN (R)	TM_K_LRR	AT4G08850	Le rep like kin pro
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.I00656	09	13607921	13610024	Singleton	<i>C. austroafricana</i>	DOWN (R) DOWN (R)	TM_K_LRR	AT4G08850	Le rep like

					<i>L. invasa</i>	UP (S)			kin pro
					<i>A. psidii</i>				
Eucgr.J02909	10	34748881	34755956	102	<i>L. invasa</i>	DOWN (R)	TM_K_LRR_MalecLike	AT2G19190	FLU rec kin
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.K02381	11	31739538	31744299	109	<i>L. invasa</i>	DOWN (R,S)	TM_K_LRR	AT1G35710	Pro far wit ric do
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.K02386	11	31822271	31826333	109	<i>C. austroafricana</i>	DOWN (S)	TM_K_LRR	AT1G35710	Pro far wit ric do
					<i>L. invasa</i>	DOWN (R)			
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.K02775	11	35174545	35176836	110	<i>C. austroafricana</i>	DOWN (S)	TM_K_LRR	AT2G31880	Le rep kin pro
					<i>L. invasa</i>	DOWN (R)			
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	DOWN			
Eucgr.L00469	scaff_24	88611	91807	-	<i>C. austroafricana</i>	DOWN (R)	TM_K_LRR	AT3G47570	Le rep kin pro
					<i>L. invasa</i>	DOWN (R)			
					<i>R. solanacearum</i>	DOWN/UP			
Eucgr.L01229	scaff_140	58951	61469	-	<i>L. invasa</i>	DOWN (R)	TM_K_LRR	AT3G47570	Le rep kin pro
					<i>A. psidii</i>	UP (S)			
					<i>R. solanacearum</i>	UP			

4. Discussion

Pattern-recognition receptors (PRRs) form the first cellular barrier against pathogens. Breeding strategies that leverage PRRs can deliver broad, non race-specific defence, in contrast to many R genes. Although several tools exist to annotate disease-resistance genes and proteins, the limited conservation of PRR domains and the use of different algorithms cause each method to recover a different gene set. Here, two state-of-the-art PRR predictors (DRAGO3 and RRGPredictor) were compared with a broad, multi-species curation from Ngou et al. (2022) in *Eucalyptus grandis*. DRAGO3 predicted 2,793 PRR-encoding genes (~ 6.1% of the proteome), RRGPredictor predicted 1,261 (~ 2.8%), and the *Eucalyptus* subset in Ngou et al. comprised 810 (~ 1.8%). This spread reflects divergent domain requirements and decision rules across strategies. After harmonising gene identifiers, categories and metadata, 730 high-confidence PRR gene candidates were identified (i.e., consistently by all three approaches), representing ~ 1.62% of protein-coding genes. Relative to *Populus trichocarpa* (~ 1% PRRs; Ngou et al., 2022), the higher proportion in *E. grandis* is consistent with high duplication rates reported for *Eucalyptus* and may reflect lineage-specific expansion (Myburg et al., 2014; Borthakur et al., 2022).

PRR classification also proved challenging because each tool presented different categories, sometimes using the same label with different composition. For example, between the PRRs assigned by DRAGO3, the broad “KIN” class showed limited overlap with PRRs predicted by RRGPredictor or by Ngou et al. (2022). Inspection of motif composition indicated that many members lacked ectodomains and therefore fall outside PRRs *sensu stricto* (Ngou et al., 2022; Restrepo-Montoya et al., 2020). The ‘UNKNOWN’ labels used for almost all PRRs by RRGPredictor likewise underscored the limits in fine-grained categorisation. To reduce tool-specific bias, classification was reframed based on PFAM domain combinations. This approach allowed categories to be tracked across methods while avoiding circularity. Eleven domain-based categories (plus a small bin without PFAM annotation) were defined. TM-K-LRR (RLK) sequences were the most abundant, whereas TM-LRR (RLP) sequences were less represented, likely reflecting detection difficulty rather than true biological scarcity (Silva et al., 2022). Many categories contained malectin or malectin-like ectodomains. Phylogenetic patterns were concordant with domain architecture: RLKs formed large clades, RLPs clustered separately, and malectin versus malectin-like ectodomains resolved into related but distinct lineages.

Within the set of 730 robust PRRs, genes were unevenly distributed across the genome. Chromosome 6 showed the highest number and proportion of PRRs and the densest clustering, suggesting an immune-enriched region. Prior work in *E. grandis* has also highlighted chromosome 6 as stress-responsive, as it

harbours the largest set of VQ (valine–glutamine) genes, regulators implicated in SA- and JA-mediated defence signalling. This convergence of VQ factors and dense PRR clustering suggests coordinated regulation of immune responses on this chromosome (Yan et al., 2019; Jing & Lin, 2015; Glazebrook, 2005). Clusters enriched for RLKs with leucine-rich repeat (LRR) domains were also observed on chromosomes 6, 8, 10, and 11, a pattern consistent with local duplication and coordinated evolution of cell-wall–sensing modules; similar proximity has been described in *Arabidopsis* (Yang et al., 2021). Duplication emerged as a major driver of PRR content: extensive tandem duplication (318 genes in 127 tandems) co-occurred with segmental, proximal, and dispersed events. Ka/Ks estimates for segmental pairs indicated predominantly purifying selection, with a minority of more divergent pairs consistent with sub- or neofunctionalisation. A great number (40.5%) of intergenic-distance clusters (< 200 kb) showed total identity with MCScanX-defined tandems (135 genes), connecting local genome structure to PRR expansion. By contrast, chromosome 5 lacked duplicated PRRs, in agreement with earlier reports of ETI-related NBS-LRR enrichment on that chromosome (Christie et al., 2015), suggesting that PTI and ETI components occupy partially distinct genomic regions.

An initial functional view based on the *E. grandis* v2.0 GO annotation in Phytozone suggested limited defence enrichment and risked under-interpretation. Re-annotation with PANNZER improved functional resolution and recovered defence-related Biological Process and Molecular Function terms, with strong signals for phosphorylation/kinase activity, signalling-receptor activity, and ATP binding, features expected for PRR pathways (Naidoo et al., 2014; rönén & Holm, 2022; Bolger et al., 2018). Given that genome annotations evolve as tools and biological knowledge improve, periodic re-annotation is warranted (Bolger et al., 2018).

A meta-analysis of four RNA-seq studies spanning *E. grandis*, *E. urophylla*, and *E. grandis* × *E. camaldulensis* under rust, canker, bacterial wilt, and gall-wasp challenge identified 283 of the 730 catalogue genes as differentially expressed, indicating broad engagement of the PRR repertoire during biotic stress. It is worth noting that additional proteomics and transcriptomics studies on other fungi, including *Calonectria pseudoreteaudii* (leaf blight) and *C. austroafricana*, were identified in the literature but could not be incorporated, as raw sequencing data were not publicly available; consequently, these works were excluded from the meta-analysis despite their relevance (Chen et al., 2015; Zwart et al., 2017). From the analysed data, approximately 19% of duplicated PRRs were differentially expressed under biotic stress, linking duplication history to functional deployment. More than half of leucine-rich repeat (LRR) domain-containing RLKs were differentially expressed, consistent with an important contribution at the interface of cell-wall status, growth, and immunity (Shiu & Bleecker, 2003; Zhang et al., 2022; Ngou et al., 2024). While LRR-RLKs are expected to recognise bacterial epitopes (e.g., flagellin, EF-Tu), leucine-rich repeat (LRR) domain-containing RLKs have been linked to sensing DAMPs and chitin-like cues (Baez et al., 2022; Gandhi et al., 2024). Responses were largely pathogen-specific and often depended on species or clones, and mixed directionality (up- and down-regulation) across pathosystems suggested context-dependent tuning of PRR signalling. Several PRR-DEGs genes present in the *E. grandis* reference were only induced in response to a single pathogen. In contrast, 93 PRR-DEGs were shared by at least two pathogens, with 16 responding to three or more. At least one instance of up-regulation was observed per gene, features consistent with broad-response candidates. Promoter analyses supported these assignments: cis-elements associated with MeJA, SA, ethylene, auxin, light, and abiotic stress were common to all PRRs, and the 16 multi-pathogen PRR-DEGs were enriched for MeJA- and light-responsive motifs, with frequent auxin motifs. These regulatory patterns align with hormone crosstalk and growth–defence trade-offs during PTI, including modulation of cell-wall properties (Kaur & Pati, 2016; Baez et al., 2022; Sajjad et al., 2023; Wilson et al., 2023).

These 16 PRR-DEGs include orthologues with cross-taxon evidence of defence roles. For example, the *Arabidopsis* gene AT1G35710 (Eucgr.K02381/K02386) was detected here under fungi (*C. austroafricana*, *A. psidii*), insects (*L. invasa*), and bacteria (*R. solanacearum*). It also has been reported as up-regulated in pine resistance to nematodes and upon exposure to *Xanthomonas* outer-membrane vesicles in *Arabidopsis*, leading to reduced pathogen growth (Modesto et al., 2021; Chalupowicz et al., 2023). AT3G14840 (Eucgr.F01307) was differentially expressed here in canker, gall-wasp, and wilt challenges, in line with prior reports as a resistance-gene analogue in olive leaf spot, as an interactor of *Xanthomonas* LPS in *Arabidopsis*, and as induced by pierid oviposition (Marchese et al., 2023; Hussan et al., 2020; Little et al., 2007). AT3G47570 (Eucgr.L00469/L01229) responded here to canker, rust, gall-wasp and wilt, consistent with up-regulation during citrus greening responses and with high non-synonymous polymorphism among candidate wilt-resistance genes in pepper (Gao et al., 2023; Kang et al., 2016). Beyond expression, the 16 candidates showed a non-random genomic arrangement: 12 were located within nine clusters distributed across chromosomes 3, 6, 8, 10, and 11, while two occurred as singletons on chromosomes 2 and 9. Several clusters consisted of receptors belonging to the same category, including groups of RLKs with leucine-rich repeat (LRR) domains on chromosomes 6 and 11. It is consistent with local duplication followed by retention of functionally related modules and mirroring patterns described for wall-sensing RLKs in *Arabidopsis* (Yang et al., 2021).

The present work uses publicly available data and resources in order to generate new knowledge with potential breeding impact, contributing to giving significance to the large amount of sequencing data available and working towards open science. Their recurrence across systems, together with the multi-pathogen responses observed here and their predicted hormone response, identify the 16 PRR-DEGs as strong candidates for subsequent validation and improvement efforts, including marker development, association mapping in breeding populations, and functional tests such as ligand identification, transient expression, or CRISPR-based perturbations in *Eucalyptus* or proxy systems (Modesto et al., 2021; Chalupowicz et al., 2023; Li et al., 2019; Marchese et al., 2023; Gao et al., 2023; Kang et al., 2016).

Declarations

Funding

This work was supported by Universidad Nacional de Moreno (PICyDT UNM VII 2021).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Author Contributions

All authors contributed to the study conception and design. Data collection and analysis were performed by YEG and CVF. All authors participated in the interpretation of the results. The first draft of the manuscript was written by YEG and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data Availability

No data were generated for this study; the data analysed were publicly available and duly referenced in the manuscript.

Acknowledgments

We would like to express our gratitude to Pamela Villalba, Juan Gabriel Rivas, and Pablo Aguilera for their valuable contributions to this work.

CVF is a member of the National Research System (SNI, Uruguay), and Programa de Desarrollo de Ciencias Básicas (PEDECIBA, Uruguay).

References

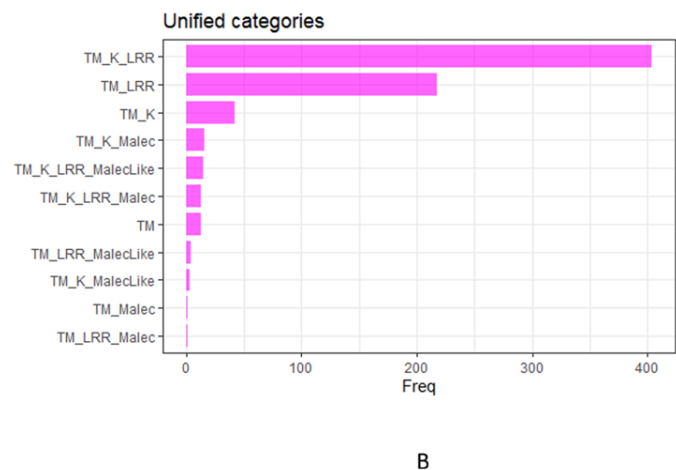
1. Alexa A, Rahnenfuhrer J (2024). topGO: Enrichment Analysis for Gene Ontology. doi:10.18129/B9.bioc.topGO, R package version 2.59.0, <https://bioconductor.org/packages/topGO>.
2. Artico, S., Ribeiro-Alves, M., Oliveira-Neto, O. B., de Macedo, L. L. P., Silveira, S., Grossi-de-Sa, M. F., Martinelli, A. P., & Alves-Ferreira, M. (2014). Transcriptome analysis of *Gossypium hirsutum* flower buds infested by cotton boll weevil (*Anthonomus grandis*) larvae. *BMC Genomics*, 15(1), 854.
3. Baez, L. A., Tichá, T., & Hamann, T. (2022). Cell wall integrity regulation across plant species. *Plant Molecular Biology*, 109(4-5), 483–504.
4. Bahar, O., Mordukhovich, G., Luu, D. D., Schwessinger, B., Daudi, A., Jehle, A. K., Felix, G., & Ronald, P. C. (2016). Bacterial Outer Membrane Vesicles Induce Plant Immune Responses. *Molecular Plant-Microbe Interactions : MPMI*, 29(5), 374–384.
5. Bartholomé, J., Mandrou, E., Mabilia, A., Jenkins, J., Nabihoudine, I., Klopp, C., Schmutz, J., Plomion, C., & Gion, J.-M. (2015). High-resolution genetic maps of *Eucalyptus* improve *Eucalyptus grandis* genome assembly. *The New Phytologist*, 206(4), 1283–1296.
6. Binagwa, P. H., Traore, S. M., Egnin, M., Bernard, G. C., Ritte, I., Mortley, D., Kamfwa, K., He, G., & Bonsi, C. (2021). Genome-Wide Identification of Powdery Mildew Resistance in Common Bean (L.). *Frontiers in Genetics*, 12, 673069.
7. Bodenhofer, U., Bonatesta, E., Horejš-Kainrath, C., & Hochreiter, S. (2015). msa: an R package for multiple sequence alignment. *Bioinformatics*, 31(24), 3997–3999.
8. Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120.
9. Bolger, M. E., Arsova, B., & Usadel, B. (2018). Plant genome and transcriptome annotations: from misconceptions to simple solutions. *Briefings in Bioinformatics*, 19(3), 437–449.
10. Borthakur, D., Busov, V., Cao, X. H., Du, Q., Gailing, O., Isik, F., Ko, J.-H., Li, C., Li, Q., Niu, S., Qu, G., Vu, T. H. G., Wang, X.-R., Wei, Z., Zhang, L., & Wei, H. (2022). Current status and trends in forest genomics. *Forestry Research*, 2, 11.
11. Calle García, J., Guadagno, A., Paytavi-Gallart, A., Saera-Vila, A., Amoroso, C. G., D'Esposito, D., Andolfo, G., Aiese Cigliano, R., Sanseverino, W., & Ercolano, M. R. (2022). PRGdb 4.0: an updated database dedicated to genes involved in plant disease resistance process. *Nucleic Acids Research*, 50(D1), D1483–D1490.
12. Cantila, A. Y., Saad, N. S. M., Amas, J. C., Edwards, D., & Batley, J. (2020). Recent Findings Unravel Genes and Genetic Factors Underlying Resistance in and Its Relatives. *International Journal of Molecular Sciences*, 22(1). <https://doi.org/10.3390/ijms22010313>
13. Chalupowicz, L., Mordukhovich, G., Assoline, N., Katsir, L., Sela, N., & Bahar, O. (2023). Bacterial outer membrane vesicles induce a transcriptional shift in arabidopsis towards immune system activation leading to suppression of pathogen growth in planta. *Journal of Extracellular Vesicles*, 12(1), e12285.
14. Charif, D., & Lobry, J. R. (2007). SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In *Structural Approaches to Sequence Evolution* (pp. 207–232). Springer Berlin Heidelberg.
15. Chen, C., Wu, Y., Li, J., Wang, X., Zeng, Z., Xu, J., Liu, Y., Feng, J., Chen, H., He, Y., & Xia, R. (2023). TBtools-II: A “one for all, all for one” bioinformatics platform for biological big-data mining. *Molecular Plant*, 16(11), 1733–1742.
16. Chen, Q., Guo, W., Feng, L., Ye, X., Xie, W., Huang, X., & Liu, J. (2015). Data for transcriptome and proteome analysis of *Eucalyptus* infected with *Calonectria pseudoreteauii*. *Data in Brief*, 3, 24–28.
17. Christie, N., Tobias, P. A., Naidoo, S., & Külheim, C. (2015). The *Eucalyptus grandis* NBS-LRR Gene Family: Physical Clustering and Expression Hotspots. *Frontiers in Plant Science*, 6, 1238.
18. Del Hierro, I., Mérida, H., Broyart, C., Santiago, J., & Molina, A. (2021). Computational prediction method to decipher receptor-glycoligand interactions in plant immunity. *The Plant Journal: For Cell and Molecular Biology*, 105(6), 1710–1726.
19. Dyda, M., Tyrka, M., Gołębiewska, G., Rapacz, M., & Wędzony, M. (2022). Genetic mapping of adult-plant resistance genes to powdery mildew in triticale. *Journal of Applied Genetics*, 63(1), 73–86.
20. Fontes, E. P. B., Santos, A. A., Luz, D. F., Waclawovsky, A. J., & Chory, J. (2004). The geminivirus nuclear shuttle protein is a virulence factor that suppresses transmembrane receptor kinase activity. *Genes & Development*, 18(20), 2545–2556.

21. Gandhi, A., Reichelt, M., Furch, A., Mithöfer, A., & Oelmüller, R. (2024). Cellooligomer/CELLOOLIGOMER RECEPTOR KINASE1 Signaling Exhibits Crosstalk with PAMP-Triggered Immune Responses and Sugar Metabolism in Arabidopsis Roots. *International Journal of Molecular Sciences*, 25(6). <https://doi.org/10.3390/ijms25063472>
22. Gao, C., Li, C., Li, Z., Liu, Y., Li, J., Guo, J., Mao, J., Fang, F., Wang, C., Deng, X., & Zheng, Z. (2023). Comparative transcriptome profiling of susceptible and tolerant citrus species at early and late stage of infection by “*Liberibacter asiaticus*.” *Frontiers in Plant Science*, 14, 1191029.
23. Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology*, 43, 205–227.
24. Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., & Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research*, 40(Database issue), D1178–D1186.
25. Gullón, P., Gullón, B., Astray, G., Munekata, P. E. S., Pateiro, M., & Lorenzo, J. M. (2020). Value-Added Compound Recovery from Invasive Forest for Biofunctional Applications: Species as a Case Study. *Molecules*, 25(18). <https://doi.org/10.3390/molecules25184227>
26. He, P., Shan, L., Lin, N.-C., Martin, G. B., Kemmerling, B., Nürnberger, T., & Sheen, J. (2006). Specific bacterial suppressors of MAMP signaling upstream of MAPKKK in Arabidopsis innate immunity. *Cell*, 125(3), 563–575.
27. Hussan, R. H., Dubery, I. A., & Piater, L. A. (2020). Identification of MAMP-responsive plasma membrane-associated proteins in Arabidopsis thaliana following challenge with different LPS chemotypes from *Xanthomonas campestris*. *Pathogens*, 9(10), 787.
28. Jing, Y., & Lin, R. (2015). The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators. *Plant Physiology*, 169(1), 371–378.
29. Jones, J. D. G., & Dangl, J. L. (2006). The plant immune system. *Nature*, 444(7117), 323–329.
30. Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A. F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., & Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30(9), 1236–1240.
31. Kang, Y. J., Ahn, Y.-K., Kim, K.-T., & Jun, T.-H. (2016). Resequencing of Capsicum annuum parental lines (YCM334 and Taeon) for the genetic analysis of bacterial wilt resistance. *BMC Plant Biology*, 16(1), 235.
32. Kaur, G., & Pati, P. K. (2016). Analysis of cis-acting regulatory elements of Respiratory burst oxidase homolog (Rboh) gene families in Arabidopsis and rice provides clues for their diverse functions. *Computational Biology and Chemistry*, 62, 104–118.
33. Krogh, A., Larsson, B., von Heijne, G., & Sonnhammer, E. L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3), 567–580.
34. Ladiges, P. Y., Udovicic, F., & Nelson, G. (2003). Australian biogeographical connections and the phylogeny of large genera in the plant family Myrtaceae. *Journal of Biogeography*, 30(7), 989–998.
35. Lescot, M., Déhais, P., Thijs, G., Marchal, K., Moreau, Y., Van de Peer, Y., Rouzé, P., & Rombauts, S. (2002). PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. *Nucleic Acids Research*, 30(1), 325–327.
36. Li, H., Dong, Z., Ma, C., Tian, X., Xiang, Z., Xia, Q., Ma, P., & Liu, W. (2019). Discovery of powdery mildew resistance gene candidates from Aegilops biuncialis chromosome 2Mb based on transcriptome sequencing. *PloS One*, 14(11), e0220089.
37. Li, J., Brader, G., & Palva, E. T. (2004). The WRKY70 transcription factor: a node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *The Plant Cell*, 16(2), 319–331.
38. Li, W., & Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13), 1658–1659.
39. Little, D., Gouhier-Darimont, C., Bruessow, F., & Reymond, P. (2007). Oviposition by pierid butterflies triggers defense responses in Arabidopsis. *Plant Physiology*, 143(2), 784–800.
40. Liu, L., Liu, J., & Xu, N. (2023). Ligand recognition and signal transduction by lectin receptor-like kinases in plant immunity. *Frontiers in Plant Science*, 14, 1201805.
41. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12), 550.
42. Lu, G., Wang, Z., Pan, Y.-B., Wu, Q., Cheng, W., Xu, F., Dai, S., Li, B., Que, Y., & Xu, L. (2023). Identification of QTLs and critical genes related to sugarcane mosaic disease resistance. *Frontiers in Plant Science*, 14, 1107314.
43. Mangwanda, R., Myburg, A. A., & Naidoo, S. (2015). Transcriptome and hormone profiling reveals Eucalyptus grandis defence responses against Chrysoperthe austroafricana. *BMC Genomics*, 16(1), 319.
44. Marchese, A., Balan, B., Trippa, D. A., Bonanno, F., Caruso, T., Imperiale, V., Marra, F. P., & Giovino, A. (2023). NGS transcriptomic analysis uncovers the possible resistance mechanisms of olive to leaf spot infection. *Frontiers in Plant Science*, 14, 1219580.
45. Mendy, B., Wang’ombe, M. W., Radakovic, Z. S., Holbein, J., Ilyas, M., Chopra, D., Holton, N., Zipfel, C., Grundler, F. M. W., & Siddique, S. (2017). Arabidopsis leucine-rich repeat receptor-like kinase NILR1 is required for induction of innate immunity to parasitic nematodes. *PLoS Pathogens*, 13(4), e1006284.
46. Minkoff, B. B., Makino, S.-I., Haruta, M., Beebe, E. T., Wrobel, R. L., Fox, B. G., & Sussman, M. R. (2017). A cell-free method for expressing and reconstituting membrane proteins enables functional characterization of the plant receptor-like protein kinase FERONIA. *The Journal of Biological Chemistry*, 292(14), 5932–5942.
47. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, 49(D1), D412–D419.

48. Modesto, I., Sterck, L., Arbona, V., Gómez-Cadenas, A., Carrasquinho, I., Van de Peer, Y., & Miguel, C. M. (2021). Insights into the mechanisms implicated in resistance to Pinewood Nematode. *Frontiers in Plant Science*, 12, 690857.
49. Myburg, A. A., Grattapaglia, D., Tuskan, G. A., Hellsten, U., Hayes, R. D., Grimwood, J., Jenkins, J., Lindquist, E., Tice, H., Bauer, D., Goodstein, D. M., Dubchak, I., Poliakov, A., Mizrahi, E., Kullán, A. R. K., Hussey, S. G., Pinard, D., van der Merwe, K., Singh, P., ... Schmutz, J. (2014). The genome of *Eucalyptus grandis*. *Nature*, 510(7505), 356–362.
50. Naidoo, S., Külheim, C., Zwart, L., Mangwanda, R., Oates, C. N., Visser, E. A., Wilken, F. E., Mamni, T. B., & Myburg, A. A. (2014). Uncovering the defence responses of *Eucalyptus* to pests and pathogens in the genomics age. *Tree Physiology*, 34(9), 931–943.
51. Navarro, L., Zipfel, C., Rowland, O., Keller, I., Robatzek, S., Boller, T., & Jones, J. D. G. (2004). The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. *Plant Physiology*, 135(2), 1113–1128.
52. Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3(5), 418–426.
53. Ngou, B. P. M., Heal, R., Wyler, M., Schmid, M. W., & Jones, J. D. G. (2022). Concerted expansion and contraction of immune receptor gene repertoires in plant genomes. *Nature Plants*, 8(10), 1146–1152.
54. Ngou, B. P. M., Wyler, M., Schmid, M. W., Kadota, Y., & Shirasu, K. (2024). Evolutionary trajectory of pattern recognition receptors in plants. *Nature Communications*, 15(1), 308.
55. Oates, C. N., Külheim, C., Myburg, A. A., Slippers, B., & Naidoo, S. (2015). The Transcriptome and Terpene Profile of *Eucalyptus grandis* Reveals Mechanisms of Defense Against the Insect Pest, *Leptocybe invasa*. *Plant & Cell Physiology*, 56(7), 1418–1428.
56. Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528.
57. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, 14(4), 417–419.
58. Restrepo-Montoya, D., Brueggeman, R., McClean, P. E., & Osorno, J. M. (2020). Computational identification of receptor-like kinases “RLK” and receptor-like proteins “RLP” in legumes. *BMC Genomics*, 21(1), 459.
59. Sajjad, M., Ahmad, A., Riaz, M. W., Hussain, Q., Yasir, M., & Lu, M.-Z. (2023). Recent genome resequencing paraded COBRA- gene family roles in abiotic stress and wood formation in Poplar. *Frontiers in Plant Science*, 14, 1242836.
60. Sakamoto, T., Deguchi, M., Brustolini, O. J. B., Santos, A. A., Silva, F. F., & Fontes, E. P. B. (2012). The tomato RLK superfamily: phylogeny and functional predictions about the role of the LRR-II-RLK subfamily in antiviral defense. *BMC Plant Biology*, 12, 229.
61. Santana Silva, R. J., & Micheli, F. (2020). RRGPredictor, a set-theory-based tool for predicting pathogen-associated molecular pattern receptors (PRRs) and resistance (R) proteins from plants. *Genomics*, 112(3), 2666–2676.
62. Santos, A. A., Carvalho, C. M., Florentino, L. H., Ramos, H. J. O., & Fontes, E. P. B. (2009). Conserved threonine residues within the A-loop of the receptor NIK differentially regulate the kinase function required for antiviral signaling. *PloS One*, 4(6), e5781.
63. Santos, S. A., Vidigal, P. M. P., Guimarães, L. M. S., Mafía, R. G., Templeton, M. D., & Alfenas, A. C. (2020). Transcriptome analysis of *Eucalyptus grandis* genotypes reveals constitutive overexpression of genes related to rust (*Austropuccinia psidii*) resistance. *Plant Molecular Biology*, 104(4-5), 339–357.
64. Shiu, S. H., & Bleeker, A. B. (2003). Expansion of the receptor-like kinase/Pelle gene family and receptor-like proteins in *Arabidopsis*. *Plant Physiology*, 132(2), 530–543.
65. Sievers, F., & Higgins, D. G. (2021). The Clustal Omega Multiple Alignment Package. *Methods in Molecular Biology*, 2231, 3–16.
66. Silva, J. C. F., Ferreira, M. A., Carvalho, T. F. M., Silva, F. F., de A Silveira, S., Brommonschenkel, S. H., & Fontes, E. P. B. (2022). RLPredictiOme, a Machine Learning-Derived Method for High-Throughput Prediction of Plant Receptor-like Proteins, Reveals Novel Classes of Transmembrane Receptors. *International Journal of Molecular Sciences*, 23(20). <https://doi.org/10.3390/ijms232012176>
67. Soneson, C., Love, M. I., & Robinson, M. D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research*, 4, 1521.
68. Takahashi, T., Murano, T., & Ishikawa, A. (2018). SOBIR1 and AGB1 independently contribute to nonhost resistance to *Pyricularia oryzae* (syn. *Magnaporthe oryzae*) in *Arabidopsis thaliana*. *Bioscience, Biotechnology, and Biochemistry*, 82(11), 1922–1930.
69. Takahashi, T., Shibuya, H., & Ishikawa, A. (2016). SOBIR1 contributes to non-host resistance to *Magnaporthe oryzae* in *Arabidopsis*. *Bioscience, Biotechnology, and Biochemistry*, 80(8), 1577–1579.
70. Thornhill, A. H., Crisp, M. D., Külheim, C., Lam, K. E., Nelson, L. A., Yeates, D. K., & Miller, J. T. (2019). A dated molecular perspective of eucalypt taxonomy, evolution and diversification. *Australian Systematic Botany*, 32(1), 29–48.
71. Törönen, P., & Holm, L. (2022). PANNZER-A practical tool for protein function prediction. *Protein Science: A Publication of the Protein Society*, 31(1), 118–128.
72. Van Bel, M., Silvestri, F., Weitz, E. M., Kreft, L., Botzki, A., Coppens, F., & Vandepoele, K. (2022). PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Research*, 50(D1), D1468–D1474.
73. Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., Lee, T.-H., Jin, H., Marler, B., Guo, H., Kissinger, J. C., & Paterson, A. H. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Research*, 40(7), e49.
74. Whitham, S. A., Quan, S., Chang, H.-S., Cooper, B., Estes, B., Zhu, T., Wang, X., & Hou, Y.-M. (2003). Diverse RNA viruses elicit the expression of common sets of genes in susceptible *Arabidopsis thaliana* plants. *The Plant Journal : For Cell and Molecular Biology*, 33(2), 271–283.
75. Wickham H. (2016). ggplot2: Elegant graphics for data analysis. New York: Springer; 2016.

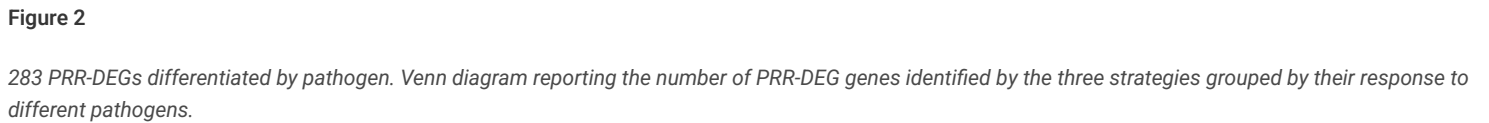
- ## Footnotes

- ## Figures



A Categories identified for putative PRR genes for each of the three strategies. Categories described according to DRAGO (pink) C, CK, CL, CLECRK, CLK, CLYSK, CT, KIN, LEC, LECRK, LYK, LYP, LYS, RLK, RLP, T, TL, TRAN, UNDEFINED. Categories according to Ngou (green) RLK-LRR -(I a XVI), RLK-LysM,, RLP-LRR, RLP-LysM. According to RRGPredictor (light blue) MLO, RLK, RLK-GNKG2, RLP, T, UNKNOWN.

B Frequency of putative PRR genes classified under the proposed new categories. TM: transmembrane, K: kinase, LRR: Leucine-Rich Repeat, Malec: malectin, MalecLike: malectin like.



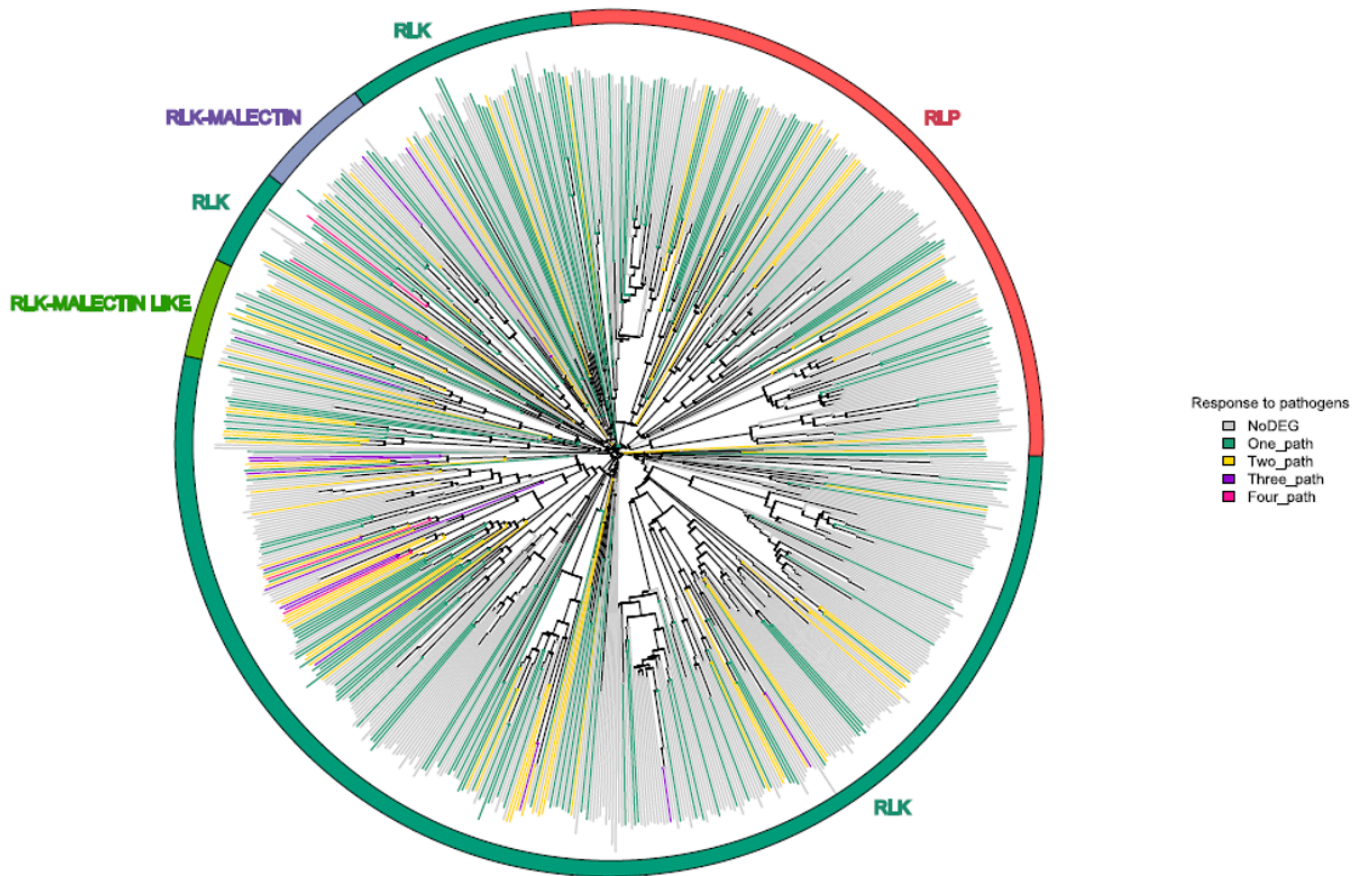


Figure 3

Phylogenetic tree for the 283 proteins encoded by PRR-DEGs. The branches of the tree are coloured in response or not to pathogens. In grey correspond the proteins identified by the three strategies that were not DEGs, in green the proteins encoded by PRR-DEGs in response to one pathogen, in yellow the proteins encoded by PRR-DEGs in response to two pathogens, in violet the proteins encoded by PRR-DEGs in response to three pathogens and in fuchsia the proteins encoded by PRR-DEGs in response to four pathogens. Clustering based on the categories for the encoded proteins is indicated above the tree.

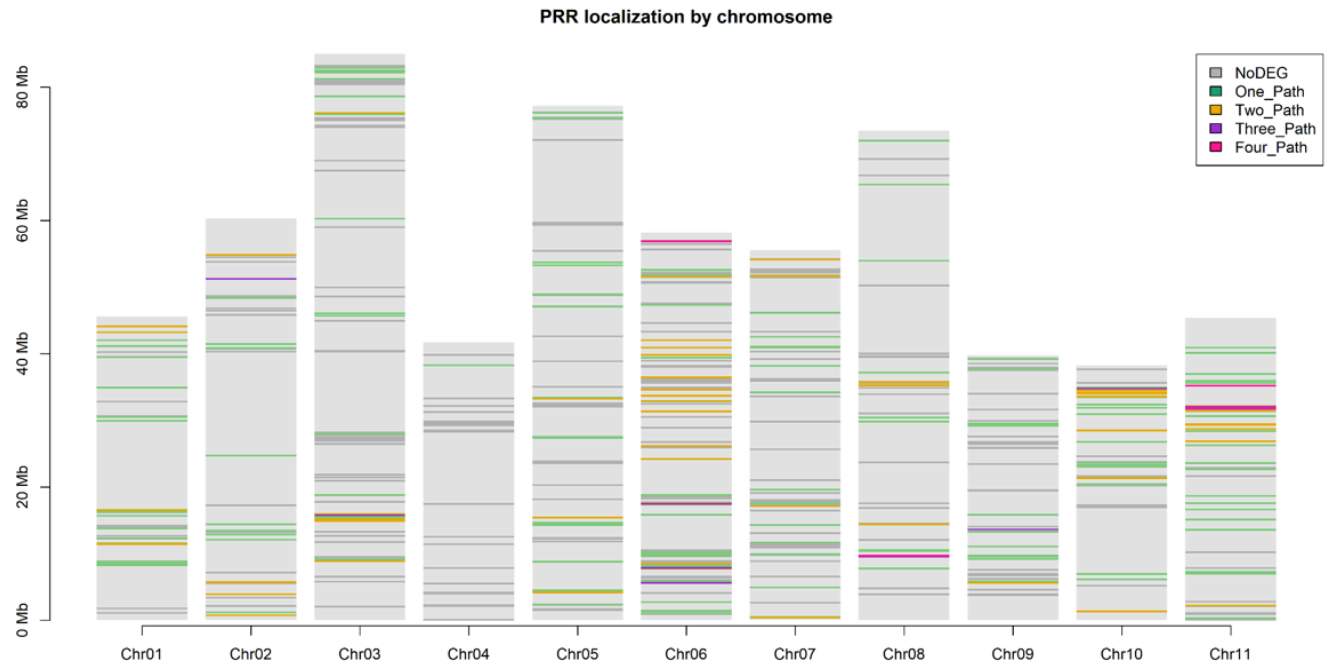


Figure 4

Localisation and chromosomal distribution of PRR and PRR-DEG in *E. grandis*. In grey PRR no DEG (to date); in green PRR-DEG in response to one pathogen, in orange PRR-DEG in response to two pathogens, in violet PRR-DEG in response to three pathogens and in fuchsia PRR-DEG in response to 4 pathogens.

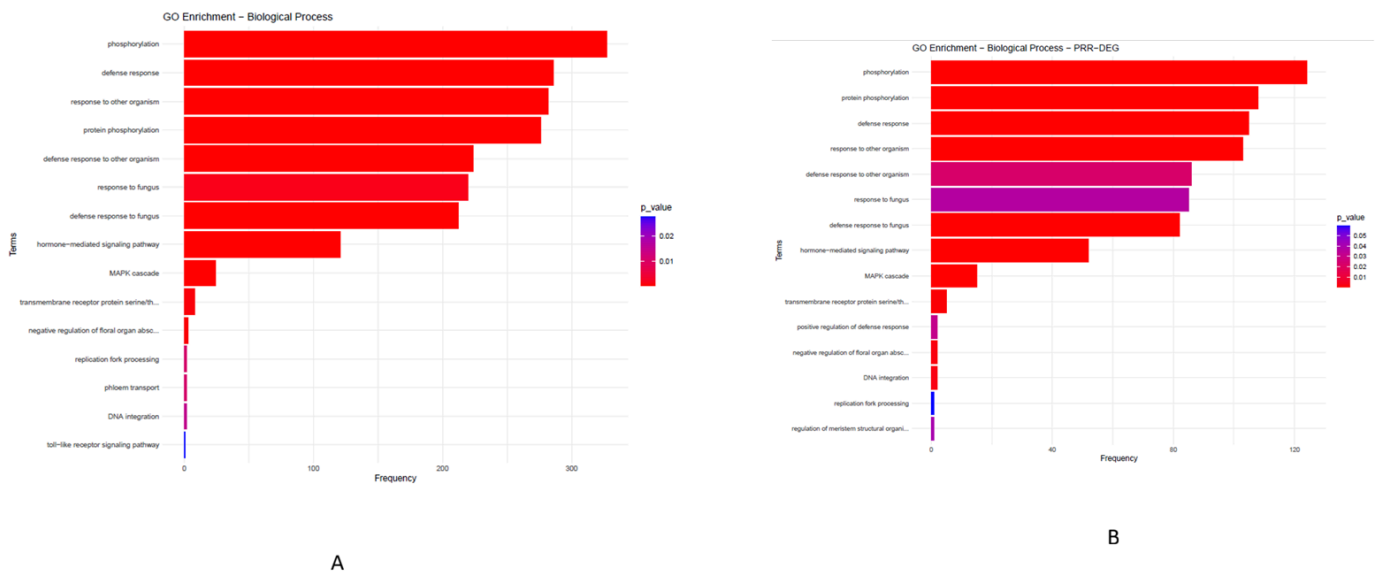


Figure 5

A New annotation GO (Biological Processes) enrichment for the 730 PRR genes identified by the three strategies. The p-value scale corresponds from $1E-30$ to 0.02793.

B New annotation GO Enrichment (Biological Processes) for the 283 PRR-DEGs. The p-value scale corresponds from $1E-30$ to 0.0597.

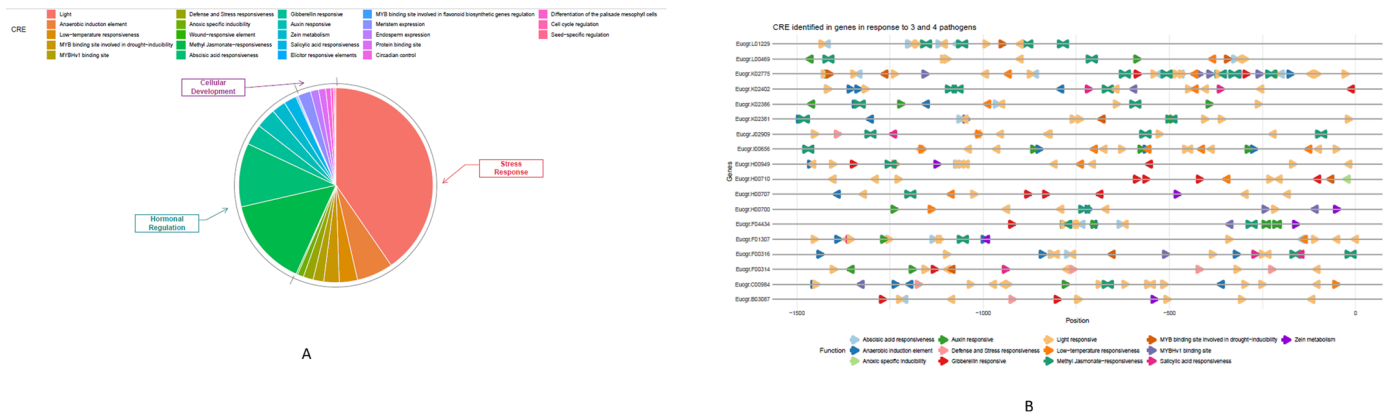


Figure 6

A Pie chart of motifs identified in *E. grandis*, based on their biological functions.

B Predicted *cis*-acting elements in the promoter regions (1500 bp upstream) of 18 PRR-DEGs that showed evidence of biotic stresses for 3 and 4 pathogens. Arrows pointing to the right indicate that the element is located on the positive strand, while arrows pointing to the left indicate its presence on the negative strand.

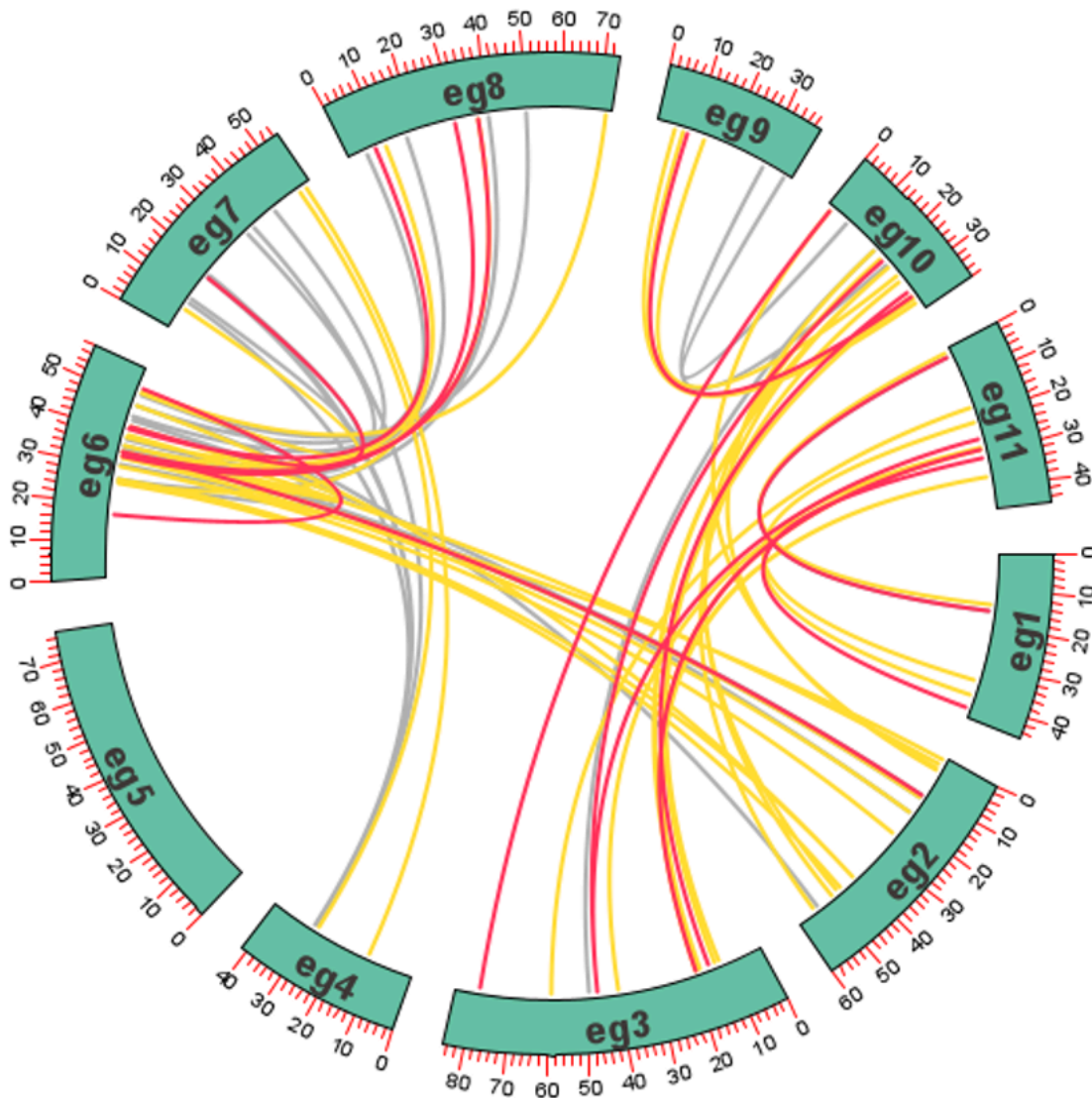


Figure 7

Page 18/19

Chromosomal location and colinearity of PRR genes in E. grandis. Blue grey lines indicate duplicated gene pairs that were not DEGs, yellow lines indicate gene pairs where at least one of them is DEG and in fuchsia duplicated pairs where both were DEGs. Eg1 - 11 refers to the chromosome number of E. grandis.

Supplementary Files

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

- [SupplementaryGonzalezetal.docx](#)
- [SI1.xlsx](#)