Identification and expression analysis of putative *NLR* genes in avocado, *Persea americana*, during infection by the oomycete, *Phytophthora cinnamomi*

by

Alicia Fick

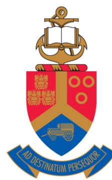
Submitted in partial fulfilment of the requirements of the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences
University of Pretoria
Pretoria

30 November 2021

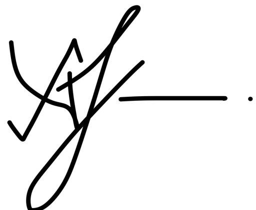
Supervisor: Prof Noëlani van den Berg
Co-supervisor: Dr Velushka Swart



UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA
Denkleiers • Leading Minds • Dikgopolo tša Dihlalefi

Declaration

I, Alicia Fick, declare that the thesis, which I hereby submit for the degree Magister Scientiae at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: 

DATE: 29 November 2021

Acknowledgements

The completion of this project would not have been possible without the following people and institutions:

To the Hans Merensky Foundation, thank you for providing the funding which made this project possible.

To my supervisor, Prof Noëlani van den Berg. Thank you for your unending kindness and support. Thank you for believing in me and encouraging me to be a better scientist. Thank you for being understanding and being a mom away from home.

To Dr Velushka Swart and Dr Robert Backer, thank you for all your insights and guidance. Our discussions regarding research have truly changed the way I think about science, and you both encouraged me to always think critically about every aspect of my work.

To my fellow lab friends, thank you for creating such a wonderful work environment. I am truly honored to be part of this wonderful team.

To my family, Carolynne, Koos and Karel, I honestly could not thank you enough for encouraging me and giving me the opportunity to pursue a degree in science. Thank you for always being interested in my work, even if you do not always understand exactly what is being said.

To my grandmother, Brenda Koster, you believed in me when I doubted myself. Thank you for being my biggest supporter every step of my journey.

To my best friend, Bernard Olivier, thank you for just being you. Thank you for always providing the humor during the hardest of days. I honestly strive to be as carefree as you are.

To Dr Jan de Vos and Lize de Vos, I could not thank you enough for your unending kindness and support, especially during the Covid-19 pandemic.

And lastly, to Monique Muller, thank you for always being there for me. Thank you for your support, love, patience, and kind heart. Thank you for always motivating me and keeping me focused.

These words would never be able to describe my true appreciation to any one of you. Thank you, from the bottom of my heart.

Preface

Avocado production in South Africa is significantly affected by the pathogen *Phytophthora cinnamomi*. This pathogen causes root necrosis in infected rootstocks, which may ultimately lead to plant death and thus economic losses. The complex molecular interactions between avocado and *P. cinnamomi* are not well understood and might be the key to breeding *P. cinnamomi* resistant rootstocks. Understanding the molecular mechanisms behind successful immune responses against *P. cinnamomi* is thus of utmost importance. Nucleotide binding-Leucine rich repeat (NLR) proteins play an important role in activating the Effector triggered immune (ETI) response after pathogen effector recognition. This immune response leads to a Hypersensitive response and cell death, which restricts *P. cinnamomi* growth during the pathogen's biotrophic stage. Furthermore, previous studies have shown that differential expression patterns of *NLR* genes influences the amplitude of the ETI response. The importance of NLR proteins during *P. cinnamomi* infection and whether they play a role in rootstock resistance levels remain unclear.

Chapter 1 of this thesis provides a detailed review of NLR proteins. The structure of these proteins is described, together with the molecular mechanisms which regulate the expression of *NLR* genes. Various examples are given regarding how *NLR* gene expression changes after pathogen infection in other plant species, and how *NLR* gene sequence variations within a plant species influence resistance levels towards various pathogens. Furthermore, methods for *NLR* identification and characterization using a genomic approach are discussed. Lastly, a summary is given on *NLR* sequences found in woody tree species and how this knowledge can be applied to avocado trees.

Chapter 2 reports on the *PaNLRs* identified in avocado using a bioinformatic approach and the VC75 avocado genome. Gene clusters were identified, and phylogenetic analysis was performed to investigate possible *PaNLR* gene duplication events. *PaNLR* expression analysis was done using dual RNA-seq data obtained from a partially resistant and susceptible rootstock inoculated with *P. cinnamomi*. The expression data was used to indicate which PaNLRs are potentially important for the detection of *P. cinnamomi* infection, and to assess whether a difference in *PaNLR* expression was evident between the two rootstocks.

Summary

Persea americana Mill is an economically important crop plant in many countries, including South Africa. South African avocado industry losses are immense due to the pathogen *Phytophthora cinnamomi* Rands, since a limited number of control methods are available to control the pathogen. Investigating *P. americana*-*P. cinnamomi* interactions on a molecular level, as well as comparing these interactions between different rootstocks, will help understand how avocado is able to suppress pathogen growth. Pathogen effectors are mainly recognized by Resistance proteins, of which Nucleotide binding-Leucine rich repeat (NLR) proteins form the majority. Differences in *NLR* expression has often been associated with resistance level differences in many plant species, including potato, tomato, and rice. NLR proteins have never been studied in *P. americana* during *P. cinnamomi* infection and may play a critical role in rootstock immunity against *P. cinnamomi.*

This study identified 161 complete *P. americana NLR* genes using a bioinformatic approach. NLR proteins were putatively identified, which may indicate which pathogen effectors they recognize. The chromosomal location of the *PaNLR* sequences were used to identify 13 gene clusters, indicating that these sequences may have originated from duplication events. Phylogenetic analysis further indicated that *PaNLR* genes within the same gene cluster have high sequence similarity, contributing to the hypothesis that gene clusters arose through gene duplication events. Phylogenetic analysis further showed that *PaNLR* sequences have little sequence similarity with *NLRs* identified in the Stout camphor tree, a close relative to *P. americana*. Expression analysis revealed that the expression of 84 *PaNLR* genes were significantly increased in response to *P. cinnamomi* inoculation, in both the partially resistant and susceptible rootstock at 6 hours post-inoculation. Increased *PaNLR* expression was sustained in the partially resistant rootstock for 24 hours after *P. cinnamomi* inoculation. However, *PaNLR* expression levels reverted to normal levels in the susceptible rootstock 12 hours post-inoculation. These results suggest that the susceptible rootstock would be unable to activate a sufficient ETI response to suppress *P. cinnamomi* growth. The results of this study provide the foundation needed for unraveling the molecular mechanisms behind *P. americana*-*P. cinnamomi* interactions.

Table of Contents

Declaration II

Acknowledgements III

Preface IV

Summary V

List of Abbreviations VIII

Chapter 1 1

Plant *NLR* gene structure, expression and function of NLR proteins: what to expect from *Persea americana*

Introduction 2

Plant-pathogen interactions 2

R protein structure 4

NLR proteins 6

NLR gene expression 8

NLR protein activation 11

Identifying *NLR* genes using genomic approaches 14

NLR genes in tree species 15

What to expect from *Persea americana*? 17

Conclusion 18

References 20

Chapter 2 29

Partially resistant avocado rootstock Dusa® employs NLR proteins to combat *Phytophthora cinnamomi* attack

Abstract 30

Introduction 31

Materials and Methods 33

Putative *PaNLR* gene identification 33

PaNLR gene cluster identification 34

Phylogenetic analysis 34

Plant inoculation and RNA-sequencing 34

Expression analysis 35

Results 36

Putative *PaNLR* genes identified in the avocado genome 36

PaNLR gene clusters identified in the VC75 genome 37
High *PaNLR* sequence similarity within *NLR* gene clusters 38
PaNLR expression following *P. cinnamomi* inoculation 40
Discussion 44
Conflict of interest 49
Author contributions 49
Funding 49
Acknowledgements 49
Data availability statement 49
References 50

List of Abbreviations

ADP	Adenosine diphosphate
ANOVA	Analysis of variance
ATP	Adenosine triphosphate
Avr	Avirulence protein
BLAST	Basic Local Alignment Search Tool
bp	Base pairs
C-terminus	Carboxyl terminus
CC	Coiled-coil domain
cDNA	Complementary DNA
CNL	CC-NB-LRR
CC_R/C_R	CNL protein with RPW8 domain
DAMP	Damage-associated molecular pattern
DNA	Deoxyribonucleic acid
DRAGO	Disease Resistance Analysis and Gene Orthology
DRL	Disease resistance-like
DTI	DAMP triggered immunity
et al.	*Et alia* – “and others”
ETI	Effector triggered immunity
ETS	Effector triggered susceptibility
GFF	General feature format
HMMER	Hidden Markov Model
hNB-LRR	Helper NLR
hpi	Hours post-inoculation
HR	Hypersensitive response
ID	Identity
JTT	Jones-Taylor-Thornton
kb	Kilo base pairs
Log_2FC	Log_2 Fold Change
LRR	Leucine rich repeat
MAMP	Microbe associated molecular pattern
MAPK	Mitogen-activated protein kinase
Mg^{2+}	Magnesium ion
MI	Mock-inoculated
miRNA	Micro-RNA
ml	Millilitre
mRNA	Messenger RNA

MTI	MAMP triggered immunity
N_2	Nitrogen
N-terminus	Amino terminus
NADPH	Nicotinamide adenine dinucleotide phosphate
NB-LRR/NLR	Nucleotide binding-Leucine rich repeat
NB/NB-ARC	Nucleotide binding domain
NCBI	National Centre for Biotechnology Information
NMD	Nonsense-mediated decay
p-value/*p*	Statistical significance
PAMP	Pathogen associated molecular pattern
PaNLR	*Persea americana NLR*
PC00g000000	*Persea americana* gene ID
phasiRNA	Phased trans-acting siRNA
PRG	Pathogen Receptor Gene
PRR	Pattern recognition receptor
PTI	PAMP triggered immunity
QTL	Quantitative trait locus
R gene	Resistance gene
R proteins	Resistance proteins
RBOHD	Respiratory burst oxidase protein D
RGA	Resistance gene analog
RLK	Receptor-like kinase
RLP	Receptor-like protein
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
ROS	Reactive oxygen species
SA	Salicylic acid
SCT	Stout camphor tree
siRNA	Small interfering RNA
STK	Serine-threonine kinase receptor
sNB-LRR	Sensor NLR
TF	Transcription factors
TIR	Toll/interleukin-1 receptor domain
TM	Transmembrane domain
TM-CC	Transmembrane Coiled-coil receptor
TMV	Tobacco mosaic virus
TNL	TIR-NB-LRR

Chapter 1

Plant *NLR* gene structure, expression and function of NLR proteins: what to expect from *Persea americana*

Introduction

In nature, plants are constantly exposed to a variety of pathogens in the form of fungi, bacteria, and viruses. As a result, plants have evolved a two-layered immune response to recognize and survive pathogen attack (Jones and Dangl 2006). The first layer consists of Pattern recognition receptors (PRRs), anchored on the plant plasma membrane, which sense conserved molecular patterns associated with pathogens or damaged cells (Tang *et al.* 2017). PRRs activate a low amplitude immune response, able to warn off non-host adapted pathogens. The second layer of the host response consists of Resistance (R) proteins, which recognize Avirulence (Avr) effector proteins produced by pathogens to suppress host immune responses. R proteins are grouped into different classes based on protein structure, with the largest class comprised of Nucleotide binding-Leucine rich repeat (NLR) proteins. NLR proteins play an important role in activating high amplitude immune responses in plants to suppress pathogen growth and infection rates (Monteiro and Nishimura 2018). NLR proteins have been studied in many plant species including *Arabidopsis thaliana*, maize, tomato, cassava and, to a lesser extent, in tree species including *Eucalyptus grandis* and *Cinnamomum kanehirae* (Chaw *et al.* 2019; Christie *et al.* 2016; Kourelis and Van der Hoorn 2018; Lozano *et al.* 2015; Neale *et al.* 2017). These studies have shown that *NLR* genes, which encode these proteins, are controlled by a variety of complex molecular mechanisms, which ensure that the correct defence response is triggered during pathogen attack. Firstly, *NLR* gene expression is regulated at both transcriptional and translational levels (Bezerra-Neto *et al.* 2020). Secondly, NLR proteins require conformational changes to be activated, which is often controlled by dimerization reactions with other NLR proteins. And thirdly, NLR protein structure influences the protein's affinity to Avr proteins (Qi and Innes 2013). All these factors regulate the timing and intensity of defence response activation, ultimately influencing resistance levels towards invading pathogens. Since only 49 NLR proteins have been identified in *Persea americana* (avocado) to date, this literature review aims to provide a broad overview of NLR protein structure, expression and function as studied in other plant species (Pérez-Torres *et al.* 2021; Van den Berg *et al.* 2018). This will then aid in the study of avocado *NLR* genes regarding protein structure and function in avocado immune responses.

Plant-pathogen interactions

Plants have evolved a wide array of complex mechanisms to sense and react to pathogen attack, which mainly relies on biochemical and genetic signals (Knepper and Day 2010). Plant pathogens include fungi, oomycetes, bacteria, viruses, nematodes, and insects which may alter the development, growth, and reproduction of plants (Dangl and Jones 2001). The plant's ability to react to pathogen infection can be explained by a simplified two-part system of innate immunity. These two parts are mainly separated based on the location of the activated plant receptor proteins. The first defence response is elicited when surface-localized immune receptors recognize and respond

to conserved molecules either associated or released by pathogens, known as Pathogen associated molecular patterns (PAMPs) or Microbe associated molecular patterns (MAMPs) (Davis and Hahlbrock 1987). Plants also recognize Damage-associated molecular patterns (DAMPs) released by neighboring plant cells damaged during pathogen attack (Matzinger 2007). Molecular patterns bind to and are recognized by plant PRRs which are transmembrane receptor proteins, including Receptor-like kinases (RLKs) and Receptor-like proteins (RLPs) (Nicaise *et al.* 2009). The recognition of these molecular patterns activates multiple signalling cascades, leading to PAMP, MAMP or DAMP triggered immunity (PTI or MTI/DTI) (Chisholm *et al.* 2006). The PTI/MTI releases Reactive oxygen species (ROS), increases cellular calcium concentrations, activates defence related gene expression, and activates Mitogen-activated protein kinases (MAPKs), aimed to restrict pathogen colonization (Eulgem 2005). These immune responses are slow, and low in amplitude, but are often sufficient in conferring resistance towards most pathogens. However, some pathogens are host-adapted and are either not recognized by the host or are able to interfere with immune responses– therefor leading to Effector triggered susceptibility (ETS) (Jones and Dangl 2006).

The second part of plant innate immunity is known as Effector triggered immunity (ETI) (Jones and Dangl 2006). ETI is triggered when *Avirulence* (*Avr*) gene products, secreted by pathogens to induce ETS, are recognized by intracellular plant receptor proteins. Avr proteins are most often recognized by NLR receptor proteins, encoded by plant resistance (*R*) genes (Monteiro and Nishimura 2018). Many R proteins recognize specific Avr proteins, thus R proteins may be determinants of plant immune response specificity. Avr recognition results in an incompatible interaction and resistance towards the invading pathogen, but only if the ETI is successful in overcoming pathogen attack. In some cases, R proteins are unable to recognize Avr proteins, or are unable to trigger an immune response sufficient to overcome the pathogen, which results in a compatible interaction and plant susceptibility. These plant-pathogen interaction processes are often explained by the Zig-zag model (Figure 1) (Jones and Dangl 2006). The PTI and ETI collectively activate defence signalling pathways, often simultaneously, and do not respond independently of one another (Wang *et al.* 2020). A recent review by Lu and Tsuda (2020) highlighted the increasing evidence which suggests that PRR proteins and R proteins activate similar signalling cascades leading to similar immune responses, including calcium influxes and MAPK activation. However, when R proteins are activated, the amplitude of these responses are increased, which ultimately leads to a stronger immune response and increased resistance. Moreover, some proteins important for PTI signalling, including BIK1 which activates NADPH oxidase RBOHD (Respiratory burst oxidase protein D), are critical for full ETI activation and ROS production in *Arabidopsis* during *Pseudomonas syringae* infection (Yuan *et al.* 2021).

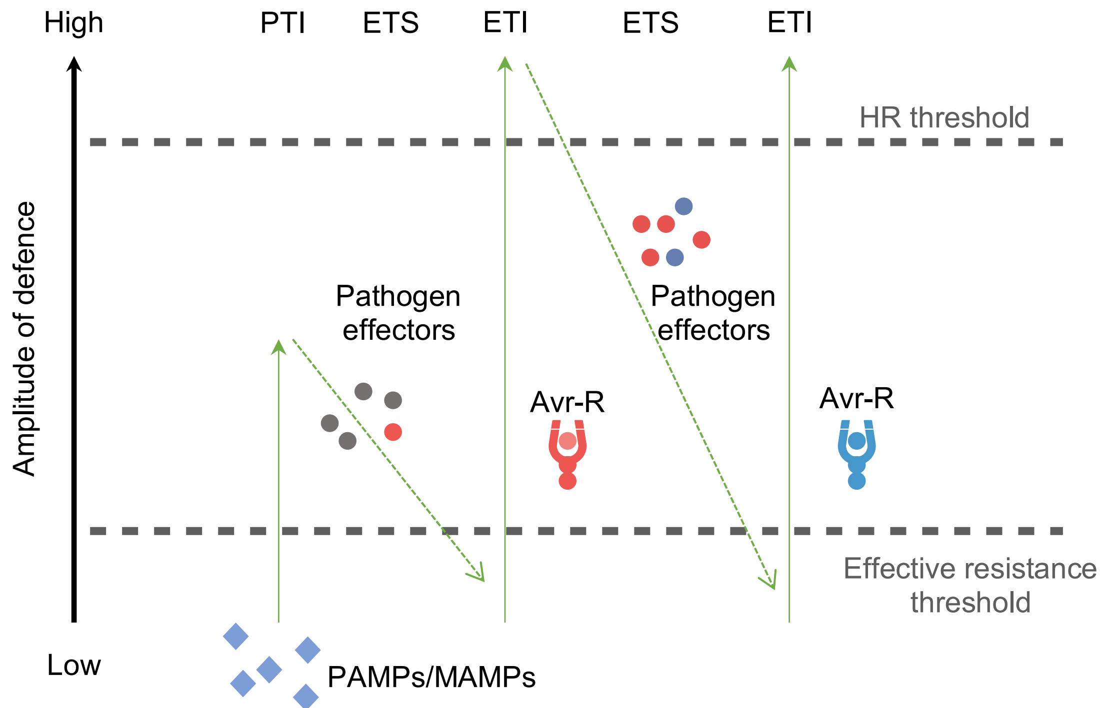


Figure 1. The Zig-zag model of plant immunity (Jones and Dangl 2006). During phase 1, plants detect Pathogen associated molecular patterns/Microbe associated molecular patterns (PAMPs/MAMPs) using Pattern recognition receptors (PRRs), which trigger PAMP triggered immunity (PTI). During phase 2, successful pathogens produce effectors such as Avirulence proteins (Avr) which interfere with the PTI response, resulting in Effector triggered susceptibility (ETS). During phase 3, one Avr protein (red) is recognized by a Resistance protein (R), triggering Effector triggered immunity (ETI). This response has a higher amplitude when compared to PTI and often induces the Hypersensitive response (HR). During phase 4, successful pathogens which have gained new Avr proteins (through horizontal gene flow or mutations) are able to evade detection, known as ETS. Phase 5 indicates the second ETI during which plants have evolved new R proteins able to recognize the new Avr proteins (blue) of the pathogen.

R protein structure

R genes encode R proteins, which are broadly grouped into five classes based on structure (Figure 2) (Bezerra-Neto *et al.* 2020). The first and main class is cytoplasmic NLR proteins, comprised of a Nucleotide binding (NB) domain located at the N-terminus of the protein, and a Leucine rich repeat (LRR) domain at the C-terminus (McDowell and Woffenden 2003). RGA2, also known as Rpi-blb1, is an NLR protein from *Solanum bulbocastanum* which confers resistance towards various *Phytophthora infestans* isolates (Van Der Vossen *et al.* 2003). The second class of R proteins, known as Receptor-like proteins (RLPs), are anchored within the plant cell membrane and have extracellular LRR domains with a short cytoplasmic C-terminus. This class of R proteins do not have a kinase domain, although it is hypothesized that these proteins interact with intracellular Receptor-like kinase proteins to allow for signal transduction (Kruijt *et al.* 2005). *Cf* genes found in tomatoes

form part of the TM-LRR class of R proteins and confer resistance to the fungus *Cladosporium fulvum* (Jones *et al.* 1994; Zhao *et al.* 2016). The third class has an extracellular LRR domain, a transmembrane domain and a Serine-threonine kinase receptor domain and is known as Receptor-like kinase proteins (RLKs) (Xu *et al.* 2000). This class of R proteins consist of more than 600 family members in *A. thaliana*, and 1100 family members in rice (*Oryza sativa*) (Shiu and Bleecker 2001). Xa21, an RLK protein in rice, confers resistance towards *Xanthomonas oryzae* pv. *oryzea* and forms part of this class (Song *et al.* 1995). The fourth class of R proteins is characterized by Serine-threonine kinase (STK) receptor domains and do not have LRR domains. The *Pto* gene found in potatoes represents this class and confers resistance to the bacterium *P. syringae* pv. *tomato* (Song *et al.* 1997). These proteins have active cytoplasmic kinases which phosphorylate serine and threonine residues. Eleven subdomains and 15 amino acid variations are often found in this R protein class; however, the phosphorylation regions are conserved (Hanks *et al.* 1988). The fifth class, known as Transmembrane Coiled-coil proteins (TM-CCs), is the most divergent and is represented by Coiled-coil (CC) domains anchored in the membrane (Kobe and Kajava 2001). RPW8 proteins in *A. thaliana* confer resistance towards *Golovinomyces* spp. fungi and form part of this R protein class (Xiao *et al.* 2001). These proteins activate the Hypersensitive response (HR) through the use of Salicylic acid and Enhanced disease susceptibility1-dependant signalling pathways (Xiao *et al.* 2003).

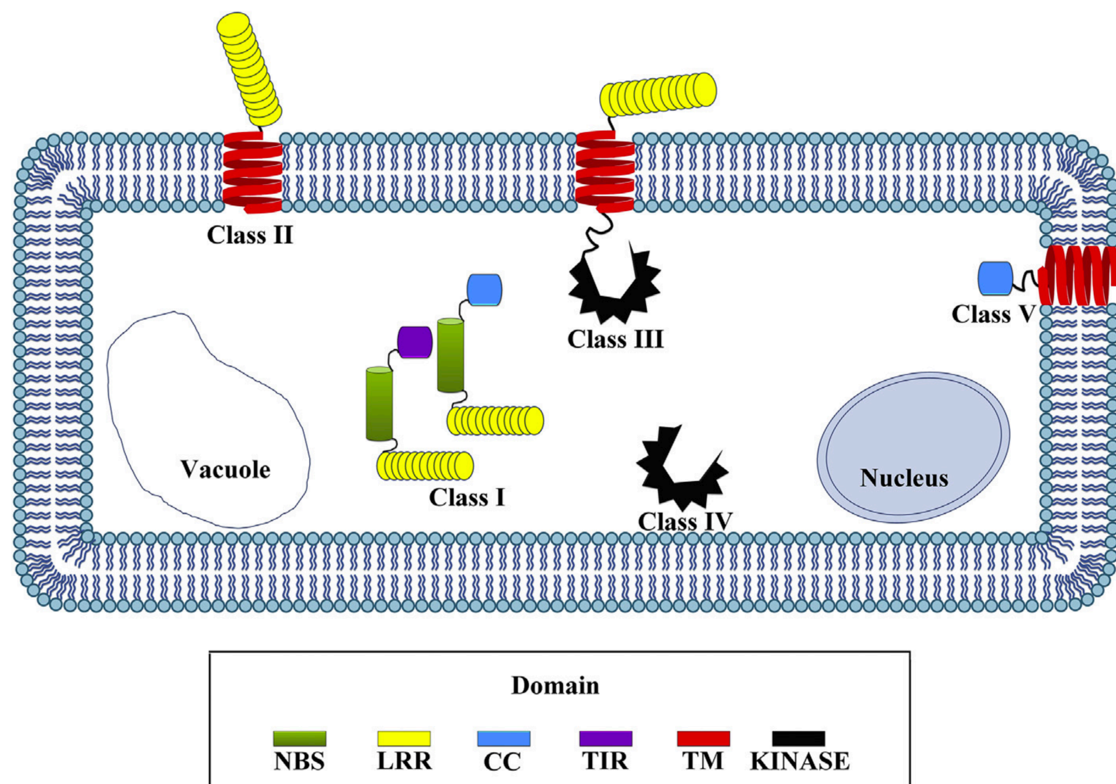


Figure 2. The five classes of plant Resistance (R) protein structure and location (Bezerra-Neto *et al.* 2020). Class I: Nucleotide binding-Leucine rich repeat proteins; Class II: Receptor-like proteins; Class III: Receptor-like kinase proteins; Class IV: Serine-threonine kinase proteins; and Class V: Transmembrane Coiled-coil proteins (CC – Coiled-coil; LRR – Leucine rich repeat; NB – Nucleotide binding; TIR – Toll/interleukin-1 receptor; TM – Transmembrane domain).

NLR proteins

NLR protein structure (Figure 3) consists of three distinct domains: a variable amino terminus domain, an NB domain in the central part of the protein (± 300 amino acids) and the LRR positioned at the C-terminus (20-30 amino acids) (Maiti *et al.* 2014). The NB domain is responsible for ADP/ATP binding and aids in signal transduction pathways, while the LRR domain is responsible for protein-protein interactions (Kedzierski *et al.* 2004; Tameling *et al.* 2006). NLR protein specificity lies in the LRR domains, due to these domains binding to pathogen Avr proteins, either directly or indirectly. *NLR* genes form part of one of the most variable and largest gene families in plants, due to the genes rapidly evolving in the arms race with pathogen *Avr* genes (Maule *et al.* 2007). The amino terminus domain splits the NB family into two subfamilies based on whether the protein has a Toll/interleukin-1 receptor (TIR) or Coiled-coil (CC) structure, named TNLs and CNLs, respectively.

The TIR domain has a homologous structure with the mammalian IL-1 receptors and Drosophila Toll protein. This domain is 135-160 amino acids in length and contains three parts of conserved residues (Xu *et al.* 2000). This domain forms five parallel β-sheet filaments (βA-βE) and five α-helices (αA-αE) connected with loops (Bernoux *et al.* 2011). Crystallography showed that the structure of the L6 TIR domain in flax (*Linum usitatissimum*) had two monomers formed by the αD and αE helices (Bernoux *et al.* 2011). Mutations within this domain broke the monomer structure and interrupted signalling activity. The activity of the TIR domain is mediated through self-association, interaction with other TIR domains or with proteins which do not have TIR domains (Ve *et al.* 2015). Research done on *Arabidopsis* RBA1 TNLs and *P. syringae* HopBA1 effectors showed that HopBA1/RBA1 interaction causes TIR self-interaction and dimerization, ultimately activating a defence response and subsequent cell death (Nishimura *et al.* 2017).

CC motifs are characterized by repeated heptad sequences interspaced with hydrophobic amino acid residues 120-200 amino acids in length (Lukasik and Takken 2009). A supercoil forms as a result of the interaction between two or more α-helices connected by a short loop. Many CC domain motifs, however, do not have a Coiled-coil structure as predicted by programmes including Basic Local Alignment Search Tool (BLAST) and Hidden Markov Model (HMMER). Motifs present in most characterized CC domains, except for the EDVID motif, follow a conserved heptad $abcdefg_n$ repeat (Mazourek *et al.* 2009). The ‘a’ and ‘d’ amino acids are mostly hydrophobic, and the ‘e’ and ‘g’ are polar and charged, resulting in amphipathic helices with both hydrophobic– and polar parts. Two helices interact due to hydrophobic interactions which gives the protein its Coiled-coil structure (Grigoryan and Keating 2008).

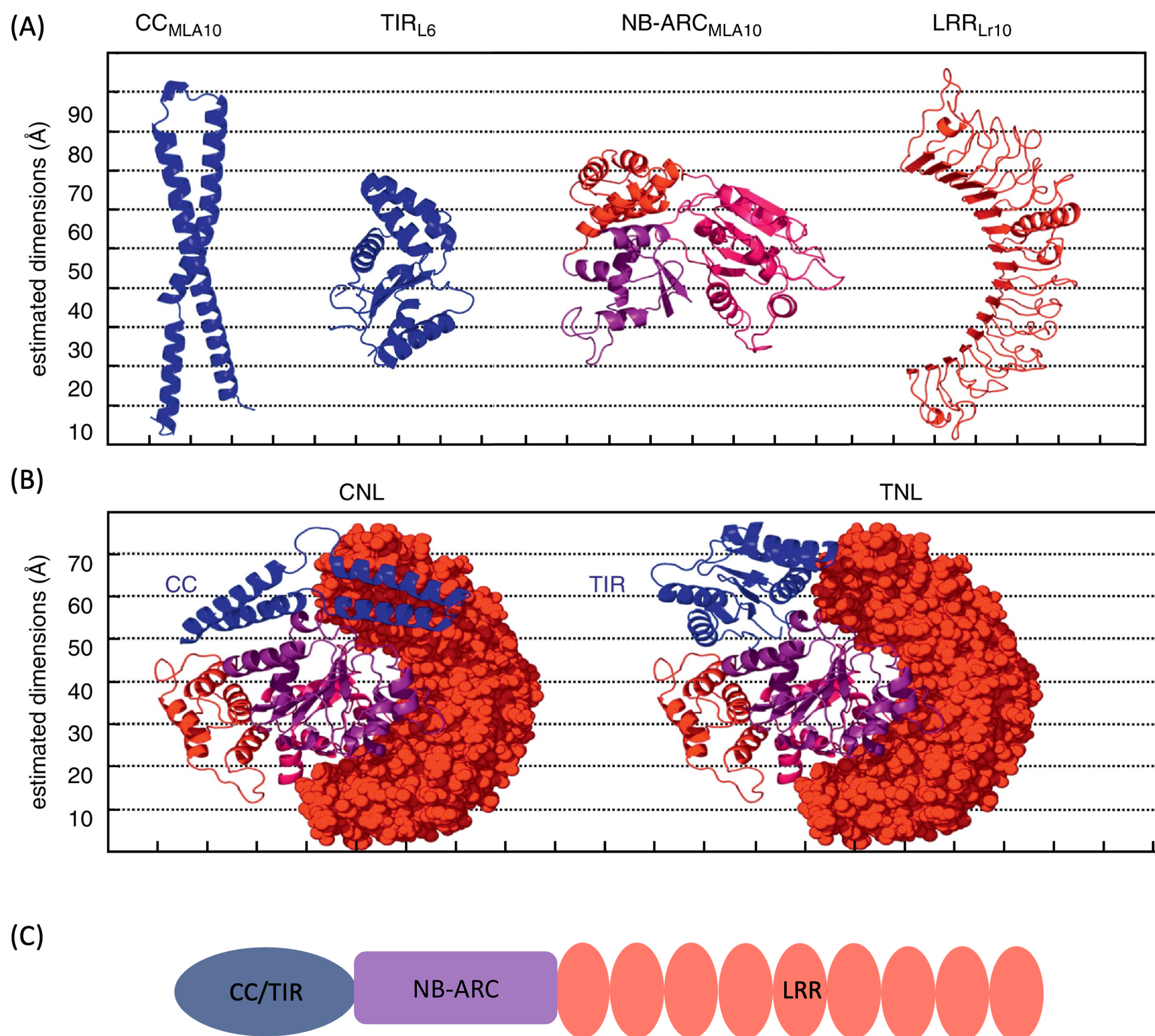


Figure 3. Structural domain models of NLR proteins. (A) Coiled-coil (CC) and Toll/interleukin-1 receptor (TIR), Nucleotide binding (NB-ARC) and Leucine rich repeat (LRR) structural domains with estimated dimensions (Å). The CC_{ML10} and TIR_{L6} models are based on crystal structures, while NB-ARC and LRR models are based on prediction models. (B) Hypothetical models of CNLs and TNLs, which show possible domain orientation and position in an inactive state. (C) Schematic 2D representation of NLR proteins with subdomains represented by different colours (Takken and Goverse 2012).

The NB domain, also called the NB-ARC domain, is hypothesized to function as a switch for molecular cascades during defence signalling (Takken *et al.* 2006). Hydrolyzed ATP is able to bind to the functional NB structure of this domain, activating the protein to trigger an immune response. The NB domain can be grouped into five highly conserved signature motifs, all of which contribute to nucleotide binding and/or hydrolysis. These motifs include the kinase-2 and kinase-3/RNBS-B motifs, the P-loop, the GxL motif and the MHDV motif (Maiti *et al.* 2014). The P-loop allows the NLR protein to bind to nucleotides, as the flexible loop of strand 1 and helix 1 accommodates the phosphate group to bind (Pathak *et al.* 2014). The consensus sequence for the P-loop follows a GxxxxGK(S/T) pattern, where x is any amino acid. The lysine amino acid found in the motif binds to

β and γ phosphates of nucleotides, with the serine/threonine amino acids binding to Mg^{2+} ions (Leipe *et al.* 2004). Mutations within these motifs ended in loss of protein function, which highlights the importance of these amino acids (Tameling *et al.* 2002). The kinase 2 motif is hypothesized to be an important regulator for the hydrolysis of ATP. This motif follows a hhhD(D/E) pattern where h represents any hydrophobic amino acid. The aspartate directs the Mg^{2+} ion required for phosphate transfer reactions, and the second amino acid is used as a catalyst during ATP hydrolysis (Tameling *et al.* 2002). Mutations within the kinase 2 motif still allow for protein interaction, although ATP hydrolysis does not occur (Tameling *et al.* 2006).

Most plant NLR proteins contain leucine rich regions which are repeated between 8 to 25 times within the C-terminus segment (Maiti *et al.* 2014). LRR domains mostly contain 25 to 30 amino acid motifs of leucine residues, or other aliphatic residues (following a LxxLxLxxN/CxL pattern) which is characteristic of proteins associated with protein-protein interactions (Kobe and Kajava 2001). This repeat motif allows the domain to have alternately arranged β-sheets, separated by a loop variable in length, and parallel α-helices forming a concave structure known as a "horseshoe" (Kobe and Kajava 2001). Most structural variations within the LRR domain are found clustered around the β-sheet, implicated to form part of the protein interaction surface. Hence, the β-sheet region is important in Avr protein recognition specificity and reflects the importance of determining resistance specificity (Qi and Innes 2013). Using genetic engineering, NLR protein specificity and affinity to Avr proteins can be increased in cases of direct Avr recognition. A single mutation within the LRR domain sequence of the rice *TNL Pikp* gene increased the binding affinity of the protein towards a *Magnaporthe oryzae* Avr protein *in vitro* and *in vivo* (De la Concepcion *et al.* 2019). The LRR domain is also responsible for regulating NLR protein activity, as these proteins become self-activating after the domain is deleted (Ade *et al.* 2007).

***NLR* gene expression**

NLR proteins form part of ETI, during which Avr pathogen proteins are recognized either directly or indirectly by the host. NLR proteins play an important role in plant defence, and their expression is tightly regulated to balance functional importance and energy expenditure (Bezerra-Neto *et al.* 2020). The presence of NLR proteins is necessary to 'scan' the cellular space for invading plant pathogens and trigger an early defence response. Over-expression of *NLR* genes can be costly and can even cause plant death or stunted growth (Tian *et al.* 2003). Thus, during this balancing act, the expression of *NLR* genes can follow two patterns: (1) low constitutive expression, such as the *Dm3* gene in lettuce (Shen *et al.* 2002), or (2) differential induction/repression of expression using epigenetics, transcription factors, alternative splicing or small RNAs as a result of applied stressors (Bezerra-Neto *et al.* 2020). Differential *NLR* gene expression levels can be seen in different species genotypes, which can ultimately influence resistance levels towards pathogens. In basket willow

trees (*Salix viminalis*) an *NLR* gene (*RGA1*) showed higher expression levels in a *Melampsora larici-epitea* resistant genotype when compared to the susceptible genotype (Martin *et al.* 2016). Furthermore, six *NLR* genes were significantly upregulated after *Chrysoporthe austroafricana* infection in resistant *E. grandis* trees only (Christie *et al.* 2016).

Epigenetic control of *NLR* gene expression was first hinted at when a *SNC1* trans-gene produced much higher levels of transcripts when compared to endogenous *SNC1* copies in *A. thaliana* (Li *et al.* 2007). This suggested that the *SNC1* gene was located within a repressive chromatin region where the promoter was under epigenetic control. Xia *et al.* (2013) identified that trimethylation of lysine 4 of histone H3 (H3K4me3) was one possible epigenetic modification which controls the expression of *SNC1*. Studies done on *A. thaliana* with recessive DNA demethylase mutations in the *ROS1*, *DML2* and *DML3* genes showed increased susceptibility to *Fusarium oxysporum* (Le *et al.* 2014). This research showed that 348 genes were differentially expressed in triple mutant (*ros1*, *dml2* and *dml3*) plants, indicating that DNA demethylases regulate or maintain the expression of defence-related genes important for resistance against *F. oxysporum*. This was later confirmed by Kong *et al.* (2020) when looking at DNA methylation patterns of *Arabidopsis NLR* genes. Ubiquitylation of histones also regulates defence gene transcription. Ubiquitin ligases, HUB1 and HUB2, mono-ubiquitylates histone H2B and H2Bub1 in *Arabidopsis* during a defence response, which increases the expression of *SNC1* and *RPP4* defence genes (Zou *et al.* 2014).

Cytosine methylation is another epigenetic mechanism found to control *NLRs*, regulated by the histone mark H3K9me2 (Lai and Eulgem 2018). The intricate interplay of methylation, methylation maintenance and demethylation controls the methylation patterns of nucleobase cytosine at position 5. Methylated cytosines (5mC) occur at CHG and GC symmetrical motifs, or CHH asymmetrical sites in plants (H representing any nucleobase except for G) (Lai and Eulgem 2018). Defence responses in *Arabidopsis* are known to affect 5mC levels globally (Deleris *et al.* 2016). For example, Avr proteins produced by *P. syringae* pv. *tomato* DC3000 induce chromatin de-condensation and hypomethylation at centromeric and pericentromeric repeats in tomato (Pavet *et al.* 2006).

Unique sequences were found within the promoter region of *NLR* genes of *Arabidopsis*, highlighting the role of *cis*-regulatory elements during a defence response (Ramkumar *et al.* 2014). Transcription factors (TFs), which are known to regulate defence-related genes, are members of the ERF, WRKY and TGA-bZIP families (Euglem 2005). WRKY TFs mainly bind to pathogen/elicitor-responsive W box *cis*-elements with hexameric consensus sequences (TTGACC/T) (Lai and Eulgem 2018). Mohr *et al.* (2010) found that *Arabidopsis NLR* gene promoters are enriched with W boxes, suggesting WRKY TFs regulate *NLR* genes. W boxes have been shown to play key roles in controlling constitutive expression and transient accumulation of defence related *RPP8* transcripts in *Arabidopsis* (Mohr *et al.* 2010). *RPP8* genes encode CNL proteins in *Arabidopsis* accession Ler, and

trigger a strong race-specific defence response during *Hyaloperonospora arabidopsidis* infection. *RPP8* showed low basal expression levels before infection, which increased in response to defence-inducing stimuli. Mutations within the *RPP8* W box caused *RPP8*-mediated defence responses to fail and resulted in susceptibility. It is unlikely that *NLRs* are exclusively regulated by WRKY TFs, however no other TF types are known with the same function (Lai and Eulgem 2018).

Alternative splicing can be used to regulate the expression of *NLR* genes on a translational level, as alternative splicing patterns can alter protein structure and change the location of stop codons within mRNA sequences. Different protein isoforms have been observed in TNLs from the tobacco *N* gene and tomato *Bs4* genes (Schornack *et al.* 2004). The tobacco *N* gene encodes two different TNL proteins, N_S and N_L, controlled by alternative splicing patterns. N_S, translated from a full-length transcript, comprises a TNL protein and is expressed before Tobacco mosaic virus (TMV) infection, until 3 hours post infection. N_L is a truncated protein lacking 13 of 14 LRR repeats and is expressed 4-8 hours after TMV infection (Dinesh-Kumar and Baker 2000). These two proteins offer full resistance to TMV, but mutations inhibiting alternative splicing result in decreased resistance, as N_L cannot be produced. Retained intronic sequences have also been reported in plants which result from alternative splicing patterns (Reddy *et al.* 2013). This form of alternative splicing can create premature in-frame stop codons, which can cause Nonsense-mediated decay (NMD) of *NLR* gene mRNA. In the case of TNLs, some mRNA molecules remain stable and are not degraded, but are translated into truncated proteins partially lacking LRR, NB-LRR or other C-terminus domains (Gassmann *et al.* 1999). These truncated proteins may play a critical role during defence activation as transcript isoforms seem to be important for resistance (Lai and Euglem 2018).

Micro-RNAs (miRNAs) and small interfering RNAs (siRNA) also have the ability to regulate gene expression at a post-transcriptional level (Axtell 2013). miRNAs have been shown to negatively regulate *NLRs* through targeting *NLR* mRNAs in angiosperms and gymnosperms, including apples, cotton, and rice (Lian *et al.* 2016; Ma *et al.* 2014; Zhu *et al.* 2014). Conserved members of the miR482/miR2118 superfamily in cotton, with variable sequences, target *NLRs* through binding to P-loop and LRR motifs (Zhu *et al.* 2014). siRNAs in *Arabidopsis*, encoded by overlapping sense and antisense transcripts within the *RPP5* cluster, can suppress *NLR* mRNA translation (Yi and Richards 2007). Recent studies have also shown that highly abundant miRNAs in legumes target 74 *NLRs* to produce phased trans-acting siRNAs (phasiRNA), which then silence 324 of 540 *NLR* genes (Zhai *et al.* 2011). This cascade is silenced in plants infected with bacterial and viral pathogens, resulting in increased *NLR* transcripts and enhanced defence states. Overexpression of two miR482-type miRNAs (nta-miR6019 and nta-miR6020) which produce phasiRNAs in *N. benthamiana,* increased the cleavage of the TMV resistance gene and decreased TMV resistance (Li *et al.* 2012). Together, these observations suggest that miRNAs serve as master regulators through producing phasiRNAs, collectively suppressing multiple *NLRs*.

The significance of differential *NLR* gene expression levels has been observed in studies done on pathogen resistant– and susceptible plants within a species. In *E. grandis*, differential *NLR* gene expression levels were studied in susceptible and resistant genotypes during *Leptocybe invasa* and *C. austroafricana* infection (Christie *et al.* 2016). The results indicated that 218 and 343 *NLR* genes were expressed differentially during *L. invasa* and *C. austroafricana* infection, respectively. Moreover, two partial *NLRs* were upregulated in resistant genotypes but downregulated in susceptible genotypes. The authors hypothesized that these partial *NLR* genes may have contributed to regulatory processes through inhibiting functional NLR protein heterodimers. Hence, microproteins with a single domain may target NLR proteins, having a dominant negative regulatory effect. Another study done on pepper plants (*Piper nigrum*), focused on the expression levels of five *NLR* genes in resistant and susceptible genotypes during *Phytophthora capsici* infection (Umadevi and Anandaraj 2017). The study showed significant differences in expression levels of the five *NLR* genes between the resistant and susceptible genotypes, during which three *NLR* genes (*PnCNBS2*, *PnCNBS3* and *PnCNBS5*) were downregulated in the resistant genotype during early infection. The results of these studies show the importance of tightly regulated *NLR* gene transcription processes, which ultimately influences resistance levels towards plant pathogens.

NLR protein activation

In the absence of a pathogen, NLR proteins are inactivated– during which the LRR domain exerts a negative role through stabilizing the ADP-bound state (Takken *et al.* 2006). *In vitro* studies using tomato showed that this protein state is highly stable, suggesting that this is the resting state (Tameling *et al.* 2006). The protein has a closed conformational form during the resting state, caused by interactions between LRR, CC and TIR domains with the NB domain (Takken and Goverse 2012) (Figure 4). When an Avr protein binds to the LRR domain during pathogen invasion, a conformational change is induced in the NB domain producing an open configuration (Lolle *et al.* 2020). ADP is then exchanged for ATP, leading to a second conformational change in the N-terminus domain, which triggers the protein's signalling potential. The protein returns to a resting state after the ATPase activity interrupts the signalling response (Tameling *et al.* 2006). NLR protein activation results in a variety of cellular changes such as increased calcium ion concentrations, increased ROS, transcriptional reprogramming, and the activation of MAPKs and HR (Cui *et al.* 2015). All these responses are aimed to reduce and inhibit pathogen growth.

NLR proteins recognize Avr proteins through three different mechanisms. The Gene-for-gene hypothesis explains how Avr proteins are recognized directly through the physical binding of the Avr protein to the NLR protein (Figure 4. A) (Flor 1971). This interaction is only possible when a corresponding *NLR* gene and *Avr* gene is present in the plant and pathogen, respectively. If either of these gene products are absent, the plant is unable to detect the pathogen which results in the

development of disease symptoms. The Gene-for-gene hypothesis explains that a specific *Avr* gene ligand binds directly to a corresponding *NLR* gene product, which results in the activation of pathogenesis related genes to ultimately induce a defence response (Gabriel and Rolfe 1990).

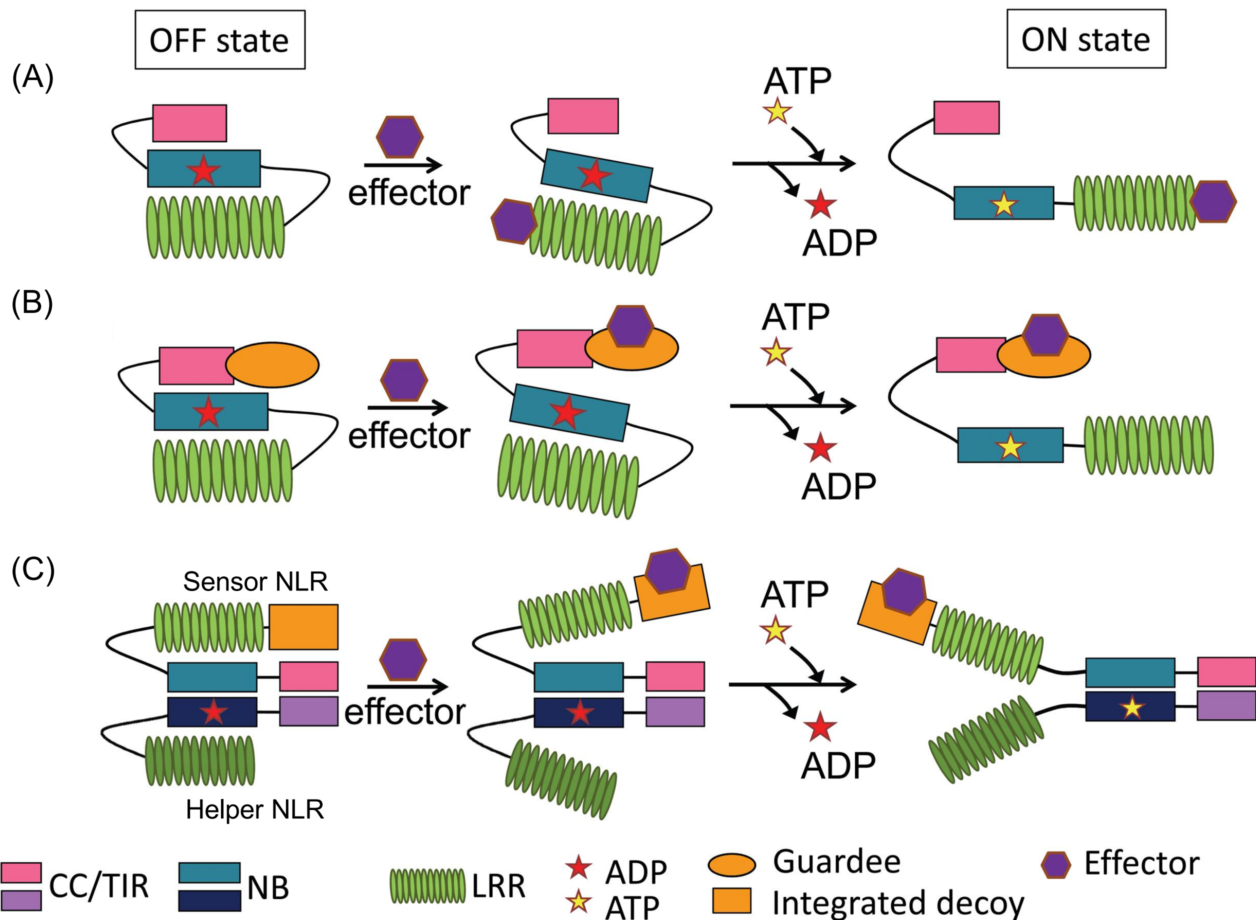


Figure 4. Models of NLR protein recognition of pathogen effectors. (A) Gene-for-gene hypothesis: Direct recognition of pathogen effector (Avirulence) proteins through binding to CC or TIR domains of Nucleotide binding-Leucine rich repeat (NLR) proteins. This is followed by the exchange of ADP with ATP, which activates the protein and downstream immune signalling. (B) Guard hypothesis: Pathogen effector proteins are recognized by cellular proteins (Guardee) associated with NLR proteins (Guard). This process leads to protein activation through the binding of ATP, and the subsequent activation of defence responses. (C) Effector proteins may be recognized through binding to a NLR dimer pair. One NLR protein with an integrated decoy domain acts as a sensor protein, able to bind and recognize an effector protein. The second NLR protein, known as the helper NLR, activates a defence response after sensing structural modifications of the sensor NLR induced by the effector. ADP is exchanged for ATP in the helper NLR, and is not required for sensor NLR function (ADP – Adenosine diphosphate; ATP – Adenosine triphosphate; CC – Coiled-coil; LRR – Leucine rich repeat; NB – Nucleotide binding; TIR – Toll/interleukin-1 receptor) (Chiang and Coaker 2015).

The Guard hypothesis is used to explain indirect protein activation, during which the Avr binds to a protein associated with a NLR protein (Figure 4. B) (Van der Biezen and Jones 1998). This model shows that NLR proteins within a plant control one or more defence proteins that are able to activate a large defence response against multiple pathogens (Bonas and Lahaye 2002). The indirect activation of NLR proteins (guard) occurs when Avr proteins bind to an associated host cellular protein (guardee). The Avr protein structurally modifies the guardee, which is detected by the guard protein through biochemical changes, ultimately triggering a defence response (Van der Biezen and Jones 1998).

Lastly, Avr proteins can be recognized when binding to a NLR pair, known as a NLR dimer (Figure 4. C). One NLR protein acts as a receptor which binds to Avr proteins either directly or indirectly, and the second activates an immune response. These NLR proteins are termed sensor (sNB-LRR) and helper (hNB-LRR) NLR, respectively (Bonardi *et al.* 2011). sNB-LRRs often possess an integrated decoy domain which mimics an Avr protein target (Césari *et al.* 2014a). Both NLR proteins are essential for proper protein function. Dimers which comprise of the same NLR proteins are known as homodimers, and dimers formed between two different NLR proteins are termed heterodimers. In rice, RGA4 and RGA5 form homo– and heterodimers through CC domain interactions. Before pathogen infection, RGA5 suppresses RGA4 activity and acts as a sNB-LRR. After Avr recognition, the dimer pair disassociates and RGA4 activates cell death (Césari *et al.* 2014b). It was suggested that NLRs with CC domains may form part of a complex network controlling ETI responses, since CNL heterodimers are found more frequently when compared to CNL homodimers (Wróblewski *et al.* 2018). However, CNL homodimers do not always have the ability to trigger an immune response compared to CNL monomers. This suggests that the oligomerization of CC domains is not required for the activation of an immune response.

After the recognition of Avr proteins, the dimerization of some NLR proteins is crucial for their function. Research has shown that *NLR* genes are often arranged as a head-to-head pair on chromosomes (Williams *et al.* 2014). This allows for increased chances of protein dimerization when proteins are expressed at the same time. An increasing amount of evidence has also shown that many NLR proteins function in pairs to activate ETI defence responses (Bentham *et al.* 2018; Swiderski *et al.* 2009). These protein pairs may be present in either a homo– or heterodimerization form of N-terminus regions. A barley CNL protein, MLA, forms a homodimer after CC domain dimerization, which is necessary for NLR protein activity (Maekawa *et al.* 2011). Heterodimers are formed in rice after CC domains of RGA4 and RGA5 NLR proteins dimerize, which can induce cell death during *M. oryzae* infection (Césari *et al.* 2014b). Moreover, several studies using TIR crystal structures have shown that two α-helices within the TIR domains of two NLR proteins are conserved in structure, allowing for these two proteins to form a dimer upon activation (Bernoux *et al.* 2011; Hyun *et al.* 2016). The interaction of TIR domains within DSC1 and WRKY19 proteins in *Arabidopsis*

contributes towards basal immunity during *Meloidogyne incognita* infection (Warmerdam *et al.* 2020). Mutations within TIR domains abolished dimerization and disrupted ETI signalling (Williams *et al.* 2014). Conformational changes within the protein upon activation may also expose potential binding surfaces, allowing for protein-protein interactions (Takken and Goverse 2012).

NLR proteins are present in many different subcellular locations, including the plasma membrane, the cytosol, and the nucleus (Qi and Innes 2013). After activation, nuclear localization, and direct interactions between some NLR proteins and transcription factors, such as WRKY, is required for activating downstream immune signalling (Wiermer *et al.* 2007). WRKY TFs are mainly associated with reprogramming gene expression patterns during the innate immune responses of plants (Eulgem and Somssich 2007). Direct interactions with TFs allow for a rapid immune response after Avr recognition, as this signalling mechanism has a short signalling pathway. One NLR protein, Rx1, has also been shown to interact directly with DNA nucleotides in response to immune activation in potato (Fenyk *et al.* 2015). However, not all NLR proteins interact directly with TFs or translocate to the nucleus. Interestingly, the RRS1-R TNL protein has a C-terminus WRKY domain, which suggests a biological role within the nucleus. This protein co-localizes in the nucleus with its cognate effector, PopP2, after *Ralstonia solanacearum* infection in *Arabidopsis* (Deslandes *et al.* 2003). How NLR proteins located in different subcellular locations trigger similar immune responses remains largely unknown (Chian and Coaker 2015).

Identifying *NLR* genes using genomic approaches

There are several studies available which utilized genomic data to identify putative *NLR* genes in different plant species (Cheng *et al.* 2017; Christie *et al.* 2016; Kim *et al.* 2017; Wang *et al.* 2019). Most of these studies utilized prediction approaches using characteristic features of *NLR* genes with databases such as InterProScan, MEME, Pfam, PROSITE and SAM (Ercolano *et al.* 2020). Through classifying proteins into families, protein function can be inferred based on the assumption that families share functional traits. This, however, requires NLR proteins that have been identified and well characterized.

NLR gene identification using genomic data includes two steps (Tirnaz *et al.* 2020). Firstly, resistant gene analog domains such as NB, LRR, CC and TIRs must be detected based on sequence similarity using different databases. Most commonly, this is done through searching for similar sequences using a query sequence in various databases. Tools such as Basic Local Alignment Search Tool (BLAST) and Hidden Markov Model (HMMER) can be used in this instance (Altschul *et al.* 1997; Finn *et al.* 2011). Specific domain sequences can be used as the consensus sequence when searching for similarity across a genome. In another approach, the genome can be used as the query sequence to search for sequences similar to domain sequences using different databases.

The Pfam database is mainly used to identify TIR, TIR_2 and LRR domains, while COILS can be used to detect CC domains (Sarris *et al.* 2016; Tirnaz *et al.* 2020). InterProscan can be used to identify protein domains through sequence analysis, using a combination of up to 14 databases (Quevillon *et al.* 2005). DRAGO (Disease Resistance Analysis and Gene Orthology) can also be used as a tool to computationally predict putative *NLR* genes (Andolfo *et al.* 2014). This tool was used to annotate more than 170 000 proteins, with the data being stored on an online database named PRG (Pathogen Receptor Gene) (http://PRGdb.org). Both well characterized and candidate plant *NLR* genes from 268 plant species can be found on this database.

The second step involves classifying candidate *NLR* genes based on domain composition (Li *et al.* 2016; Tirnaz *et al.* 2020). This is mainly done using custom scripts based on domain configuration. *NLR* gene classification is based on whether the gene encodes an NB domain. Further classification can be done based on the presence of CC, TIR or LRR domains. Genes can also be classified in "Other categories" if the genes have more than one domain. For example, the presence of both TIR and CC domains. Another example would be when genes with disordered domains, such as when a TIR domain is found at the C-terminus rather than the N-terminus. Li *et al.* (2016) recently developed the RGAugury pipeline which combines both steps of domain detection and classification. This pipeline can be used for large-scale genome-wide *NLR* gene identification.

Identifying putative *NLR* genes can be limited by several factors. Genome assembly quality can affect *NLR* gene accuracy and efficiency, for example repeat masking can influence prediction frequency (Bayer *et al.* 2018). Repeat sequences, often found in transposable elements, may contain *NLR* gene related domain sequences, which hinder the identification process. Increasing repeat masking approaches may result in true *NLR* genes being categorized as repeat sequences, but decreasing masking may result in wrongly categorized *NLR* genes (Bayer *et al.* 2018). For this reason, it is vital to evaluate masking approaches to increase the accuracy of *NLR* gene identification. Large differences in *NLR* gene numbers have been reported for the same species when different identification approaches and different parameters were used. Tools such as RGAugury greatly increase *NLR* gene identification and identification accuracy (Tirnaz *et al.* 2020).

***NLR* genes in tree species**

Our current understanding of NLR proteins mostly originated from studies done on crop plants and model species. Advances in genomic technologies and whole-genome sequencing has only recently led to the identification of NLR proteins in tree species, since large genome sizes made these processes strenuous. A broad and considerably larger *NLR* gene family has been identified in woody perennial species, using reference tree genome sequences, when compared to herbaceous species such as *Arabidopsis* and tomato (Figure 5). These tree species include *C. kanehirae*, *E. grandis*,

Hevea brasiliensis, *Picea abies*, *Pinus taeda*, *Populus trichocarpa* and *Prunus mume*, and fruit tree species *Malus* x *domestica* (apple), *Prunus persica* (peach), *Vitis vinifera* (grape), and *Ziziphus jujube* (jujube) (Chaw *et al.* 2019; Neale *et al.* 2017).

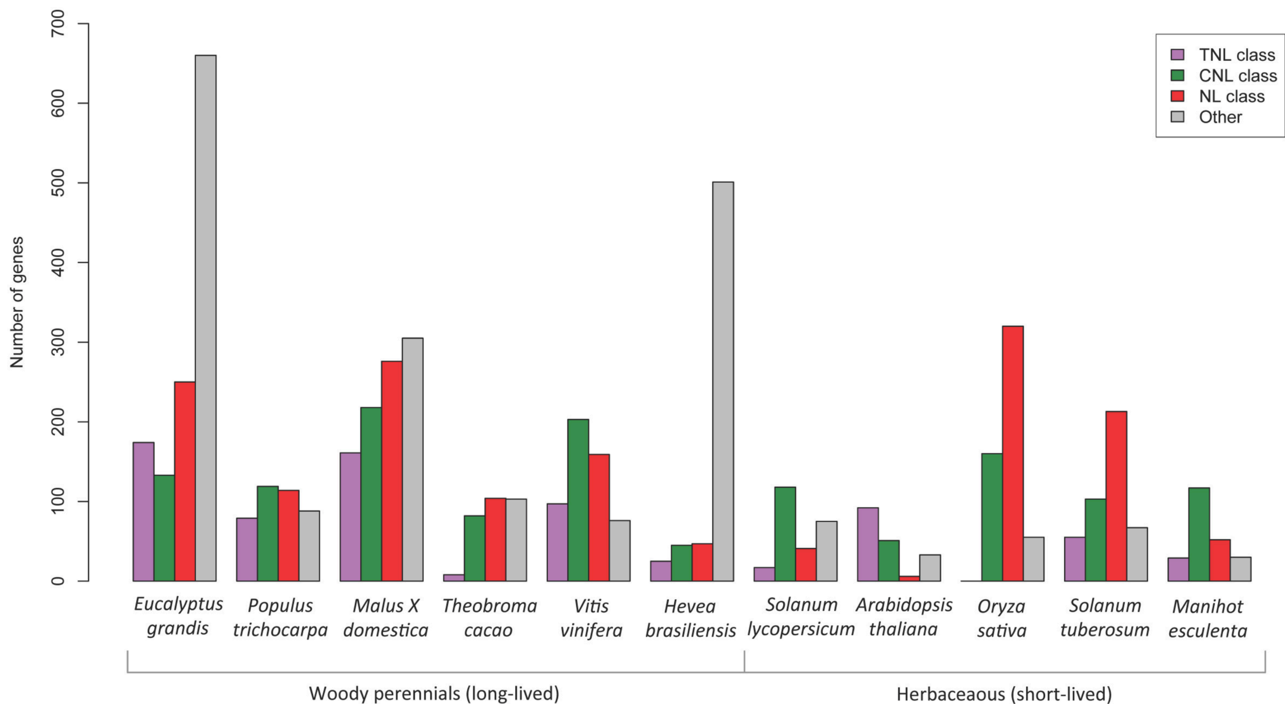


Figure 5. Comparison of putative Nucleotide binding-Leucine rich repeat (*NLR*) gene numbers between different plant species. An increased number of total *NLR* genes have been identified in woody perennial species when compared to herbaceous species. An increased number of Coiled-coil Nucleotide binding-Leucine rich repeat (*CNL*) class *NLR* genes are found in most tree species when compared to Toll/interleukin-1 receptor Nucleotide binding-Leucine rich repeat (*TNL*) class *NLR* genes, with the exception of *Eucalyptus grandis* and *Arabidopsis thaliana*. The “other” category includes *TN* and *CN* gene models lacking a Leucine rich repeat domain sequence, and *NL* gene models lacking Coiled-coil or Toll/interleukin-1 receptor domain sequences (Christie *et al.* 2016).

Tandem duplications (in *C. kanehirae*, *E. grandis*, *P. trichocarpa*, *M.* x *domestica* and *V. vinifera*), linked to *NLR* gene family expansions, may have contributed to increased plant fitness, as tree species need long-term defence strategies against pathogens (Tobias and Guest 2014). *CNLs* are often more abundant in tree species when compared to *TNLs*, contrary to the pattern seen in *Arabidopsis*. Some tree species, including *E. grandis*, *P. abies* and *P. taeda*, do not follow this pattern as certain *TNL* sequences have been expanded in angiosperms and conifer species (Neale *et al.* 2017). These duplications often result in large *NLR* gene clusters, such as seen in *E. grandis* where 136 *NLR* gene clusters were identified, containing an average of eight *NLR* genes per cluster (Christie *et al.* 2016). These clusters contain the same class of *NLR* genes (either *TNLs* or *CNLs*) more often when compared to mixed-class clusters– a trait also seen in *A. thaliana* (Leister 2004).

Functional studies on tree NLR proteins have mainly been done on *M.* x *domestica* and *V. vinifera*. Arya *et al.* (2014) identified 1015 *NLR* gene sequences using the whole-genome sequence of *M.* x *domestica*. Furthermore, the expression profile of 26 *NLR* genes were analysed after apple leaves were infected with *Alternaria alternata*, *Podosphaera leucotricha*, sap sucking insects and viruses (species unknown). This study identified *NLR* genes which may be important in activating defence responses after infection of various pathogens. Host miRNAs have also been observed to play a role in disease resistance levels in apples following *A. alternata* infection (Ma *et al.* 2014). Md-miRLn11, which is able to cleave *Md-NBS* mRNA transcripts (a *NLR* gene transcript), decreases disease resistance in a susceptible apple variety. Md-miRLn11 expression was higher in the susceptible variety when compared to a resistant variety, with the opposite trend seen for *Md-NBS.* When *Md-NBS* was expressed as a transgene in susceptible plants, disease resistance was significantly increased.

In *V. vinifera*, 142 *NLR* genes have been identified (Malacarne *et al.* 2012). When *Vitis amurensis* was infected with two different *Plasmopara viticola* stains, ZJ-1-1 and JL-7-2, disease symptoms were only observed in plants infected with ZJ-1-1 (Li *et al.* 2015). Further research showed that 37 *NLR* genes were expressed at significantly higher levels during JL-7-2 infection when compared to ZJ-1-1 infection, ultimately increasing disease resistance towards JL-7-2. This research shows that a plant species may have drastically different resistance levels towards different pathogen strains, and highlights the importance of *NLR* genes during defence response activation. *Erysiphe necator* infection of seven partially resistant and two susceptible *V. vinifera* accessions showed the importance of *NLR* gene expression regulation (Goyal *et al.* 2020). *NLR* genes within partially resistant accessions showed the same trend in expression level differences during infection, while the opposite trend was seen in the susceptible accessions. However, differences in the number of differentially expressed genes were still observed between the different accessions, ranging from nine to 23 genes.

What to expect from *Persea americana*?

Avocado is an industrially important crop tree in South Africa, with a growing popularity due to avocado fruits having high levels of monosaturated fats. *Phytophthora cinnamomi* and more recently, *Rosellinia necatrix*, are devastating pathogens impacting the avocado industry (Hardham 2005; Zumaquero *et al.* 2019). The use of partially resistant avocado rootstocks is one control method currently used to control *P. cinnamomi*. However, there are currently no commercial rootstocks resistant or partially resistant towards *R. necatrix* (Zumaquero *et al.* 2019). Understanding the molecular mechanisms behind avocado immune responses might accelerate rootstock screening programmes, and ultimately breeding programmes aimed at increasing resistance towards these pathogens.

To date, only 49 putative avocado NLR proteins have been identified using transcriptomic approaches (Péres-Torres *et al.* 2021; Van den Berg *et al.* 2018). In the study done by Pérez-Torres *et al.* (2021), *NLR* expression was investigated in Hass avocado stems infected with *Fusarium kuroshium.* In total, 48 complete *NLR* sequences were identified, with four *NLRs* showing differential expression patterns following *F. kuroshium* infection. Interestingly, only one *TNL* was identified during this study. Using transcriptomic approaches for *NLR* identification, however, limits the number of *NLRs* being identified to the number of *NLR* genes being expressed. Thus, only using transcriptomic data for *NLR* identification would prevent the identification of most *NLR* genes within a genome. Three avocado genomes have recently been sequenced by Rendón-Anaya *et al.* (2019) and the Avocado Genome Consortium (unpublished). This enables further putative *NLR* gene identification. The closest relative to *P. americana* in which a comprehensive set of *NLR* genes has been identified, was the Stout camphor tree (*C. kanehirae*), part of the Lauraceae family. In total, 387 *NLR* gene models were identified using the Stout camphor tree (SCT) genome, 317 of which belonged to the *CNL* class (Chaw *et al.* 2019). Phylogenetic analysis of these *NLR* gene sequences suggested that *NLR* clades diversified independently within eudicots, monocots and magnoliids (Chaw *et al.* 2019). Furthermore, it was found that the largest SCT *NLR* gene clades grouped closest with those of depauperate eudicot *NLR* gene clades. The large amount of SCT *NLR* gene sequences may be a result of two rounds of genome duplication within the Lauraceae family ancestry, and another genome duplication in the Magnoliaceae ancestry (Cui *et al.* 2006).

Based on these findings, it is expected that the avocado genome might harbor a large, diverse amount of *NLR* gene sequences. It is also expected that the number of *CNL* gene sequences may be larger when compared to *TNL* gene sequences, as suggested by many other tree species, including the SCT (Neale *et al.* 2017). Enriched tandem duplications in avocado were described by Rendón-Anaya *et al.* (2019), hypothesized to be important for metabolic processes and adaptations towards fungal pathogen resistance. These tandem duplications may lead to large numbers of *NLR* gene sequences in *P. americana*, as seen in the SCT genome (Chaw *et al.* 2019; Cui *et al.* 2006). As seen in *E. grandis* trees infected with *C. austroafricana* and *L. invasa* described earlier, it is expected that avocado trees will show variable expression patterns of *NLR* genes in response to different pathogen infections (Christie *et al.* 2016).

Conclusion

NLR proteins serve an important role in recognizing pathogen attack and activating plant defence responses. Many different mechanisms, including levels of gene expression, domain interactions and protein-protein interactions regulate how and when these proteins operate. These mechanisms are crucial to ensure the correct timing of defence responses and ultimately plant survival. An

increasing amount of progress has been made in understanding the molecular mechanisms underlying NLR protein control and function. A comprehensive understanding of NLR proteins is of particular interest since these proteins can be targeted and used to increase plant resistance towards pathogens. However, many gaps remain as more evidence shows an increasingly complex system. Studies focusing on *NLR* gene sequence and expression level differences between resistant and susceptible plants within a species are notably lacking. Furthermore, studies identifying NLR proteins and their respective molecular functionality within tree species, especially fruit tree species, are also lacking.

Studying these molecular mechanisms in avocado trees will aid in understanding different avocado-pathogen interactions, and which NLR proteins are important for activating defence responses in avocado trees. Moreover, such studies will further the knowledge of *NLR* gene evolution, functionality and immune signalling cascades, which can be applied to other plant species. Ultimately, knowledge in these areas may increase the success of tree screening programmes which aim to produce crop cultivars with enhanced pathogen resistance. The research study following this chapter will aim to identify putative avocado *NLR* genes using a newly sequenced avocado genome, VC75. The expression pattern of the identified *NLR* genes will also be analyzed after a partially resistant and susceptible avocado rootstock is infected with *P. cinnamomi* to identify *NLR* genes which may play a role in avocado defence responses.

References

Ade, J., B. J. DeYoung, C. Golstein and R. W. Innes, 2007 Indirect activation of a plant Nucleotide binding site-Leucine rich repeat protein by a bacterial protease. Proceedings of the National Academy of Sciences 104: 2531-2536.

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang *et al.*, 1997 Gapped BLAST and PSI-BLAST: a new generation of protein database search programmes. Nucleic Acids Research 25: 3389-3402.

Andolfo, G., F. Jupe, K. Witek, G. J. Etherington, M. R. Ercolano *et al.*, 2014 Defining the full tomato *NB-LRR* resistance gene repertoire using genomic and cDNA RenSeq. BMC Plant Biology 14: 120.

Arya, P., G. Kumar, V. Acharya and A. K. Singh, 2014 Genome-wide identification and expression analysis of NBS-encoding genes in *Malus* x *domestica* and expansion of *NBS* genes family in Rosaceae. PloS one 9: e107987-e107987.

Axtell, M. J., 2013 Classification and comparison of small RNAs from plants. Annual Review in Plant Biology 64: 137-159.

Bayer, P. E., D. Edwards and J. Batley, 2018 Bias in resistance gene prediction due to repeat masking. Nature Plants 4: 762-765.

Bentham, A. R., R. Zdrzałek, J. C. De la Concepcion and M. J. Banfield, 2018 Uncoiling CNLs: structure/function approaches to understanding CC domain function in plant NLRs. Plant and Cell Physiology 59: 2398-2408.

Bernoux, M., T. Ve, S. Williams, C. Warren, D. Hatters *et al.*, 2011 Structural and functional analysis of a plant resistance protein TIR domain reveals interfaces for self-association, signalling, and autoregulation. Cell Host and Microbe 9: 200-211.

Bezerra-Neto, J. P., F. C. Araújo, J. R. C. Ferreira-Neto, R. L. O. Silva, A. N. C. Borges *et al.*, 2020 *NBS-LRR* genes-plant health sentinels: structure, roles, evolution and biotechnological applications, pp. 63-120 in *Applied Plant Biotechnology for Improving Resistance to Biotic Stress*, edited by P. Poltronieri and Y. Hong. Academic Press, Amsterdam.

Bonardi, V., S. Tang, A. Stallmann, M. Roberts, K. Cherkis *et al.*, 2011 Expanded functions for a family of plant intracellular immune receptors beyond specific recognition of pathogen effectors. Proceedings of the National Academy of Sciences of the United States of America 108: 16463-16468.

Bonas, U., and T. Lahaye, 2002 Plant disease resistance triggered by pathogen-derived molecules: refined models of specific recognition. Current Opinion in Microbiology 5: 44-50.

Césari, S., M. Bernoux, P. Moncuquet, T. Kroj and P. Dodds, 2014a A novel conserved mechanism for plant NLR protein pairs: the ‘integrated decoy’ hypothesis. Frontiers in Plant Science 5: 606.

Césari, S., H. Kanzaki, T. Fujiwara, M. Bernoux, V. Chalvon *et al.*, 2014b The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. The EMBO Journal 33: 1941-1959.

Chaw, S. M., Y. C. Liu, Y. W. Wu, H. Y. Wang, C. Y. I. Lin *et al.*, 2019 Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nature Plants 5: 63-73.

Cheng, Y., Q. Ma, H. Ren, Q. Xia, E. Song *et al.*, 2017 Fine mapping of a *Phytophthora*-resistance gene *RpsWY* in soybean (*Glycine max* L.) by high-throughput genome-wide sequencing. Theoretical and Applied Genetics 130: 1041-1051.

Chiang, Y. H., and G. Coaker, 2015 Effector triggered immunity: NLR immune perception and downstream defence responses. The *Arabidopsis* Book 2015: e0183.

Chisholm, S. T., G. Coaker, B. Day and B. J. Staskawicz, 2006 Host-microbe interactions: shaping the evolution of the plant immune response. Cell 124: 803-814.

Christie, N., P. A. Tobias, S. Naidoo and C. Külheim, 2016 The *Eucalyptus grandis NBS-LRR* gene family: physical clustering and expression hotspots. Frontiers in Plant Science 6: 1238.

Cui, H., K. Tsuda and J. E. Parker, 2015 Effector triggered immunity: from pathogen perception to robust defence. Annual Review of Plant Biology 66: 487-511.

Cui, L., P. K. Wall, J. H. Leebens-Mack, B. G. Lindsay, D. E. Soltis *et al.*, 2006 Widespread genome duplications throughout the history of flowering plants. Genome Research 16: 738-749.

Dangl, J. L., and J. D. Jones, 2001 Plant pathogens and integrated defence responses to infection. Nature 411: 826-833.

Davis, K. R., and K. Hahlbrock, 1987 Induction of defence responses in cultured parsley cells by plant cell wall fragments. Plant Physiology 84: 1286-1290.

De la Concepcion, J. C., M. Franceschetti, D. MacLean, R. Terauchi, S. Kamoun *et al.*, 2019 Protein engineering expands the effector recognition profile of a rice NLR immune receptor. eLife 8: e47713.

Deleris, A., T. Halter and L. Navarro, 2016 DNA methylation and demethylation in plant immunity. Annual Review in Phytopathology 54: 579-603.

Deslandes, L., J. Olivier, N. Peeters, D. X. Feng, M. Khounlotham *et al.*, 2003 Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. Proceedings of the National Academy of Sciences 100: 8024-8029.

Dinesh-Kumar, S. P., and B. J. Baker, 2000 Alternatively spliced *N* resistance gene transcripts: their possible role in Tobacco mosaic virus resistance. Proceedings of the National Academy of Sciences 97: 1908-1913.

Eulgem, T., 2005 Regulation of the *Arabidopsis* defence transcriptome. Trends in Plant Science 10: 71-78.

Eulgem, T., and I. E. Somssich, 2007 Networks of WRKY transcription factors in defence signalling. Current Opinion in Plant Biology 10: 366-371.

Ercolano, M. R., G. Andolfo and L. Frusciante, 2020 Informatic tools and platforms for enhancing plant *R*-gene discovery process, pp. 121-135 in *Applied Plant Biotechnology for Improving Resistance to Biotic Stress*, edited by P. Poltronieri and Y. Hong. Academic Press, Cambridge.

Fenyk, S., P. D. Townsend, C. H. Dixon, G. B. Spies, A. De San Eustaquio Campillo *et al.*, 2015 The potato Nucleotide binding-Leucine rich repeat (NLR) immune receptor Rx1 is a pathogen-dependent DNA-deforming protein. The Journal of Biological Chemistry 290: 24945-24960.

Finn, R. D., J. Clements and S. R. Eddy, 2011 HMMER web server: interactive sequence similarity searching. Nucleic Acids Research 39: W29-W37.

Flor, H. H., 1971 Current status of the Gene-for-gene concept. Annual Review of Phytopathology 9: 275-296.

Gassmann, W., M. E. Hinsch and B. J. Staskawicz, 1999 The *Arabidopsis RPS4* bacterial-resistance gene is a member of the *TIR-NBS-LRR* family of disease-resistance genes. Plant Journal 20: 265-277.

Gabriel, D. W., and B. G. Rolfe, 1990 Working models of specific recognition in plant-microbe interactions. Annual Review of Phytopathology 28: 365-391.

Goyal, N., G. Bhatia, S. Sharma, N. Garewal, A. Upadhyay *et al.*, 2020 Genome-wide characterization revealed role of *NBS-LRR* genes during powdery mildew infection in *Vitis vinifera*. Genomics 112: 312-322.

Grigoryan, G., and A. E. Keating, 2008 Structural specificity in Coiled-coil interactions. Current Opinion is Structural Biology 18: 477-483.

Hanks, S. K., A. M. Quinn and T. Hunter, 1988 The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241: 42-52.

Hardham, A. R., 2005 *Phytophthora cinnamomi*. Molecular Plant Pathology 6: 589-604.

Hyun, K. G., Y. Lee, J. Yoon, H. Yi and J. J. Song, 2016 Crystal structure of *Arabidopsis thaliana* SNC1 TIR domain. Biochemical and Biophysical Research Communications 481: 146-152.

Jones, D. A., C. M. Thomas, K. E. Hammond-Kosack, P. J. Balint-Kurti and J. D. Jones, 1994 Isolation of the tomato *Cf-9* gene for resistance to *Cladosporium fulvum* by transposon tagging. Science 266: 789-793.

Jones, J. D. G., and J. L. Dangl, 2006 The plant immune system. Nature 444: 323-329.

Kedzierski, Ł., J. Montgomery, J. Curtis and E. Handman, 2004 Leucine rich repeats in host-pathogen interactions. Archivum Immunologiae et Therapiae Experimentalis 52: 104-112.

Kim, S., J. Park, S. I. Yeom, Y. M. Kim, E. Seo *et al.*, 2017 New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication. Genome Biology 18: 210.

Knepper, C., and B. Day, 2010 From perception to activation: the molecular-genetic and biochemical landscape of disease resistance signalling in plants. The *Arabidopsis* Book 8: e012.

Kobe, B., and A. V. Kajava, 2001 The Leucine rich repeat as a protein recognition motif. Current Opinion in Structural Biology 11: 725-732.

Kong, W., X. Xia, Q. Wang, L. W. Liu, S. Zhang *et al.*, 2020 Impact of DNA demethylases on the DNA methylation and transcription of *Arabidopsis NLR* genes. Frontiers in Genetics 11: 460.

Kourelis, J., and R. A. L. Van der Hoorn, 2018 Defended to the nines: 25 years of resistance gene cloning identifies nine mechanisms for R protein function. The Plant Cell 30: 285-299.

Kruijt, M., D. J. Kip, M. H. Joosten, B. F. Brandwagt and P. J. de Wit, 2005 The *Cf-4* and *Cf-9* resistance genes against *Cladosporium fulvum* are conserved in wild tomato species. Molecular Plant-Microbe Interactions 18: 1011-1021.

Lai, Y., and T. Eulgem, 2018 Transcript-level expression control of plant *NLR* genes. Molecular Plant Pathology 19: 1267-1281.

Le, T. N., U. Schumann, N. A. Smith, S. Tiwari, P. C. Au *et al.*, 2014 DNA demethylases target promoter transposable elements to positively regulate stress responsive genes in *Arabidopsis*. Genome Biology 15: 458.

Leipe, D. D., E. V. Koonin and L. Aravind, 2004 STAND, a class of P-loop NTPases including animal and plant regulators of programmed cell death: multiple, complex domain architectures, unusual phyletic patterns, and evolution by horizontal gene transfer. Journal of Molecular Biology 343: 1-28.

Leister, D., 2004 Tandem and segmental gene duplication and recombination in the evolution of plant disease resistance genes. Trends in Genetics 20: 116-122.

Li, F., D. Pignatta, C. Bendix, J. O. Brunkard, M. M. Cohn *et al.*, 2012 MicroRNA regulation of plant innate immune receptors. Proceedings of the National Academy of Sciences of the United States of America 109: 1790-1795.

Li, P., X. Quan, G. Jia, J. Xiao, S. Cloutier *et al.*, 2016 RGAugury: a pipeline for genome-wide prediction of Resistance gene analogs (RGAs) in plants. BMC Genomics 17: 852.

Li, Y., S. Yang, H. Yang and J. Hua, 2007 The *TIR-NB-LRR* gene *SNC1* is regulated at the transcript level by multiple factors. Molecular Plant-Microbe Interactions 20: 1449-1456.

Li, X., J. Wu, L. Yin, Y. Zhang, J. Qu *et al.*, 2015 Comparative transcriptome analysis reveals defence-related genes and pathways against downy mildew in *Vitis amurensis* grapevine. Plant Physiology and Biochemistry 95: 1-14.

Lian, S., W. K. Cho, S. M. Kim, H. Choi and K. H. Kim, 2016 Time-course small RNA profiling reveals rice miRNAs and their target genes in response to rice stripe virus infection. PLoS ONE 11: e0162319.

Lolle, S., D. Stevens and G. Coaker, 2020 Plant NLR-triggered immunity: from receptor activation to downstream signalling. Current Opinion in Immunology 62: 99-105.

Lozano, R., M. T. Hamblin, S. Prochnik and J. L. Jannink, 2015 Identification and distribution of the *NBS-LRR* gene family in the cassava genome. BMC Genomics 16: 360.

Lu, Y., and K. Tsuda, 2020 Intimate association of PRR– and NLR-mediated signalling in plant immunity. Molecular Plant-Microbe Interactions 34: 3-14.

Lukasik, E., and F. L. Takken, 2009 STANDing strong, resistance proteins instigators of plant defence. Current Opinion in Plant Biology 12: 427-436.

Ma, C., Y. Lu, S. Bai, W. Zhang, X. Duan *et al.*, 2014 Cloning and characterization of miRNAs and their targets, including a novel miRNA-targeted NBS-LRR protein class gene in apple (Golden Delicious). Molecular Plant 7: 218-230.

Maekawa, T., W. Cheng, L. N. Spiridon, A. Töller, E. Lukasik *et al.*, 2011 Coiled-coil domain-dependent homodimerization of intracellular barley immune receptors defines a minimal functional module for triggering cell death. Cell Host and Microbe 9: 187-199.

Maiti, S., J. Basak and A. Pal, 2014 Current understanding on plant *R*-genes/proteins and mechanisms of defence responses against biotic stresses, pp. 93-126 in *Review of Plant Pathology*, edited by B. N. Chakraborty. Scientific Publishers, India.

Malacarne, G., M. Perazzolli, A. Cestaro, L. Sterck, P. Fontana *et al.*, 2012 Deconstruction of the (Paleo) polyploid grapevine genome based on the analysis of transposition events involving *NBS* resistance genes. PLoS ONE 7: e29762.

Martin, T., A. C. Rönnberg-Wästljung, J. Stenlid and B. Samils, 2016 Identification of a differentially expressed *TIR-NBS-LRR* gene in a major QTL associated to leaf rust resistance in *Salix*. PLoS ONE 11: e0168776.

Matzinger, P., 2007 Friendly and dangerous signals: is the tissue in control? Nature Immunology 8: 11-13.

Maule, A. J., C. Caranta and M. I. Boulton, 2007 Sources of natural resistance to plant viruses: status and prospects. Molecular Plant Pathology 8: 223-231.

Mazourek, M., E. T. Cirulli, S. M. Collier, L. G. Landry, B. C. Kang *et al.*, 2009 The fractionated orthology of Bs2 and Rx/Gpa2 supports shared synteny of disease resistance in the Solanaceae. Genetics 182: 1351-1364.

McDowell, J. M., and B. J. Woffenden, 2003 Plant disease resistance genes: recent insights and potential applications. Trends in Biotechnology 21: 178-183.

Mohr, T. J., N. D. Mammarella, T. Hoff, B. J. Woffenden, J. G. Jelesko *et al.*, 2010 The *Arabidopsis* downy mildew resistance gene *RPP8* is induced by pathogens and Salicylic acid and is regulated by W box *cis* elements. Molecular Plant-Microbe Interactions 23: 1303-1315.

Monteiro, F., and M. T. Nishimura, 2018 Structural, functional, and genomic diversity of plant NLR proteins: an evolved resource for rational engineering of plant immunity. Annual Review of Phytopathology 56: 243-267.

Neale, D. B., P. J. Martínez-García, A. R. De La Torre, S. Montanari and X. X. Wei, 2017 Novel insights into tree biology and genome evolution as revealed through genomics. Annual Review of Plant Biology 68: 457-483.

Nicaise, V., M. Roux and C. Zipfel, 2009 Recent advances in PAMP-triggered immunity against bacteria: Pattern recognition receptors watch over and raise the alarm. Plant Physiology 150: 1638-1647.

Nishimura, M. T., R. G. Anderson, K. A. Cherkis, T. F. Law, Q. L. Liu *et al.*, 2017 TIR-only protein RBA1 recognizes a pathogen effector to regulate cell death in *Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America 114: e2053-e2062.

Pathak, E., N. Atri and R. Mishra, 2014 Analysis of P-loop and its flanking region subsequence of diverse NTPases reveals evolutionary selected residues. Bioinformation 10: 216-220.

Pavet, V., C. Quintero, N. M. Cecchini, A. L. Rosa and M. E. Alvarez, 2006 *Arabidopsis* displays centromeric DNA hypomethylation and cytological alterations of heterochromatin upon attack by *Pseudomonas syringae*. Molecular Plant-Microbe Interactions 19: 577-587.

Pérez-Torres, C. A., E. Ibarra-Laclette, E. E. Hernández-Domínguez, B. Rodríguez-Haas, A. J. Pérez-Lira *et al*., 2021 Molecular evidence of the avocado defence response to *Fusarium kuroshium* infection: a deep transcriptome analysis using RNA-seq. PeerJ 9: e11215.

Qi, D., and R. Innes, 2013 Recent advances in plant NLR structure, function, localization, and signalling. Frontiers in Immunology 4: 348.

Quevillon, E., V. Silventoinen, S. Pillai, N. Harte, N. Mulder *et al.*, 2005 InterProScan: protein domains identifier. Nucleic Acids Research 33: W116-W120.

Ramkumar, G., M. S. Madhav, A. K. Biswal, S. J. S. Rama Devi, K. Sakthivel *et al*., 2014 Genome-wide identification and characterization of transcription factor binding motifs of *NBS-LRR* genes in rice and *Arabidopsis*. Journal of Genomes and Exomes 3: 7-15.

Reddy, A. S. N., Y. Marquez, M. Kalyna and A. Barta, 2013 Complexity of the alternative splicing landscape in plants. The Plant Cell 25: 3657-3683.

Rendón-Anaya, M., E. Ibarra-Laclette, A. Méndez-Bravo, T. Lan, C. Zheng *et al.*, 2019 The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. Proceedings of the National Academy of Sciences 116: 17081.

Sarris, P. F., V. Cevik, G. Dagdas, J. D. G. Jones and K. V. Krasileva, 2016 Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. BMC Biology 14: 8.

Schornack, S., A. Ballvora, D. Gurlebeck, J. Peart, D. Baulcombe *et al.*, 2004 The tomato resistance protein Bs4 is a predicted non-nuclear TIR-NB-LRR protein that mediates defence responses to severely truncated derivatives of AvrBs4 and overexpressed AvrBs3. Plant Journal 37: 46-60.

Shen, K. A., D. B. Chin, R. Arroyo-Garcia, O. E. Ochoa, D. O. Lavelle *et al.*, 2002 *Dm3* is one member of a large constitutively expressed family of Nucleotide binding site-Leucine rich repeat encoding genes. Molecular Plant-Microbe Interactions 15: 251-261.

Shiu, S. H., and A. B. Bleecker, 2001 Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. Proceedings of the National Academy of Sciences of the United States of America 98: 10763-10768.

Song, W. Y., L. Y. Pi, G. L. Wang, J. Gardner, T. Holsten *et al.*, 1997 Evolution of the rice *Xa21* disease resistance gene family. The Plant Cell 9: 1279-1287.

Song, W. Y., G. L. Wang, L. L. Chen, H. S. Kim, L. Y. Pi *et al.*, 1995 A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science 270: 1804-1806.

Swiderski, M. R., D. Birker and J. D. G. Jones, 2009 The TIR domain of TIR-NB-LRR resistance proteins is a signalling domain involved in cell death induction. Molecular Plant-Microbe Interactions 22: 157-165.

Takken, F. L., and A. Goverse, 2012 How to build a pathogen detector: structural basis of NB-LRR function. Current Opinion in Plant Biology 15: 375-384.

Takken, F. L., M. Albrecht and W. I. Tameling, 2006 Resistance proteins: molecular switches of plant defence. Current Opinion in Plant Biology 9: 383-390.

Tameling, W. I., S. D. Elzinga, P. S. Darmin, J. H. Vossen, F. L. Takken *et al.*, 2002 The tomato *R* gene products I-2 and MI-1 are functional ATP binding proteins with ATPase activity. The Plant Cell 14: 2929-2939.

Tameling, W. I. L., J. H. Vossen, M. Albrecht, T. Lengauer, J. A. Berden *et al.*, 2006 Mutations in the NB-ARC domain of I-2 that impair ATP hydrolysis cause autoactivation. Plant Physiology 140: 1233-1245.

Tang, D., G. Wang and J. M. Zhou, 2017 Receptor kinases in plant-pathogen interactions: more than pattern recognition. The Plant Cell 29: 618-637.

Tian, D., M. B. Traw, J. Q. Chen, M. Kreitman and J. Bergelson, 2003 Fitness costs of *R* gene-mediated resistance in *Arabidopsis thaliana*. Nature 423: 74-77.

Tirnaz, S., Y. Zhang and J. Batley, 2020 Genome-wide mining of disease resistance gene analogs using conserved domains, pp. 365-375 in *Legume Genomics: Methods and Protocols*, edited by M. Jain and R. Garg. Springer US, New York.

Tobias, P. A., and D. I. Guest, 2014 Tree immunity: growing old without antibodies. Trends in Plant Science 19: 367-370.

Umadevi, P., and M. Anandaraj, 2017 Genotype specific host resistance for *Phytophthora* in black pepper (*Piper nigrum* L.). Physiological and Molecular Plant Pathology 100: 237-241.

Van den Berg, N., W. Mahomed, N. A. Olivier, V. Swart and B. G. Crampton, 2018 Transcriptome analysis of an incompatible *Persea americana-Phytophthora cinnamomi* interaction reveals the involvement of SA– and JA-pathways in a successful defence response. PLoS ONE 13: e0205705.

Van der Biezen, E. A., and J. D. G. Jones, 1998 The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. Current Biology 8: R226-R228.

Van Der Vossen, E., A. Sikkema, B. T. L. Hekkert, J. Gros, P. Stevens *et al.*, 2003 An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. The Plant Journal 36: 867-882.

Ve, T., S. J. Williams and B. Kobe, 2015 Structure and function of Toll/interleukin-1 receptor/resistance protein (TIR) domains. Apoptosis 20: 250-261.

Wang, L., L. Zhao, X. Zhang, Q. Zhang, Y. Jia *et al.*, 2019 Large-scale identification and functional analysis of *NLR* genes in blast resistance in the Tetep rice genome sequence. Proceedings of the National Academy of Sciences 116: 18479-18487.

Wang, W., B. Feng, J. M. Zhou and D. Tang, 2020 Plant immune signalling: advancing on two frontiers. Journal of Integrative Plant Biology 62: 2-24.

Warmerdam, S., M. G. Sterken, O. C. A. Sukarta, C. C. Van Schaik, M. E. P. Oortwijn *et al.*, 2020 The TIR-NB-LRR pair DSC1 and WRKY19 contributes to basal immunity of *Arabidopsis* to the root-knot nematode *Meloidogyne incognita*. BMC Plant Biology 20: 73.

Wiermer, M., K. Palma, Y. Zhang and X. Li, 2007 Should I stay or should I go? Nucleocytoplasmic trafficking in plant innate immunity. Cellular Microbiology 9: 1880-1890.

Williams, S. J., K. H. Sohn, L. Wan, M. Bernoux, P. F. Sarris *et al.*, 2014 Structural basis for assembly and function of a heterodimeric plant immune receptor. Science 344: 299-303.

Wróblewski, T., L. Spiridon, E. C. Martin, A. J. Petrescu, K. Cavanaugh *et al.*, 2018 Genome-wide functional analyses of plant Coiled-coil NLR-type pathogen receptors reveal essential roles of their N-terminal domain in oligomerization, networking, and immunity. PLoS Biology 16: e2005821.

Xia, S., Y. T. Cheng, S. Huang, J. Win, A. Soards *et al.*, 2013 Regulation of transcription of Nucleotide binding-Leucine rich repeat-encoding genes *SNC1* and *RPP4* via H3K4 trimethylation. Plant Physiology 162: 1694-1705.

Xiao, S., S. Brown, E. Patrick, C. Brearley and J. G. Turner, 2003 Enhanced transcription of the *Arabidopsis* disease resistance genes *RPW8.1* and *RPW8.2* via a Salicylic acid-dependent amplification circuit is required for hypersensitive cell death. The Plant Cell 15: 33-45.

Xiao, S., S. Ellwood, O. Calis, E. Patrick, T. Li *et al.*, 2001 Broad-spectrum mildew resistance in *Arabidopsis thaliana* mediated by RPW8. Science 291: 118-120.

Xu, Y., X. Tao, B. Shen, T. Horng, R. Medzhitov *et al.*, 2000 Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. Nature 408: 111-115.

Yi, H., and E. J. Richards, 2007 A cluster of disease resistance genes in *Arabidopsis* is coordinately regulated by transcriptional activation and RNA silencing. The Plant Cell 19: 2929-2939.

Yuan, M., Z. Jiang, G. Bi, K. Nomura, M. Liu *et al.*, 2021 Pattern-recognition receptors are required for NLR-mediated plant immunity. Nature 592: 105-109.

Zhai, J., D. H. Jeong, E. De Paoli, S. Park, B. D. Rosen *et al.*, 2011 MicroRNAs as master regulators of the plant *NB-LRR* defence gene family via the production of phased, trans-acting siRNAs. Genes and Development 25: 2540-2553.

Zhao, T., J. Jiang, G. Liu, S. He, H. Zhang *et al.*, 2016 Mapping and candidate gene screening of tomato *Cladosporium fulvum*-resistant gene *Cf-19*, based on high-throughput sequencing technology. BMC Plant Biology 16: 51.

Zhu, Q. H., L. Fan, Y. Liu, H. Xu, D. Llewellyn *et al.*, 2014 miR482 regulation of *NBS-LRR* defence genes during fungal pathogen infection in cotton. PLoS ONE 8: e84390.

Zou, B., D. L. Yang, Z. Shi, H. Dong and J. Hua, 2014 Monoubiquitination of histone 2B at the disease resistance gene locus regulates its expression and impacts immune responses in *Arabidopsis*. Plant Physiology 165: 309-318.

Zumaquero, A., E. Martínez-Ferri, A. J. Matas, B. Reeksting, N. A. Olivier *et al.*, 2019 *Rosellinia necatrix* infection induces differential gene expression between tolerant and susceptible avocado rootstocks. PLoS ONE 14: e0212359.

Chapter 2

Partially resistant avocado rootstock Dusa® employs NLR proteins to combat *Phytophthora cinnamomi* attack

Submitted to Frontiers in Plant Science

Abstract

Avocado is an important agricultural food crop in many countries worldwide. *Phytophthora cinnamomi*, a hemibiotrophic oomycete, remains one of the most devastating pathogens within the avocado industry, as it is near impossible to eradicate from areas where the pathogen is present. A key aspect to Phytophthora root rot disease management is the use of avocado rootstocks partially resistant to *P. cinnamomi*, which demonstrates an increased immune response following infection. In plant species, Nucleotide binding-Leucine rich repeat (NLR) proteins form an integral part of pathogen recognition and Effector triggered immune responses (ETI). To date, a comprehensive set of *P. americana NLR* genes have yet to be identified, though their discovery is crucial to understanding the molecular mechanisms underlying *P. americana-P. cinnamomi* interactions. In this study, a total of 161 *PaNLR* genes were identified in the *P. americana* VC75 genome. These putative resistance genes were characterized using bioinformatic approaches and grouped into 13 distinct *PaNLR* gene clusters, with phylogenetic analysis revealing high sequence similarity within these clusters. Additionally, *PaNLR* expression levels were analyzed in both a partially resistant (Dusa®) and a susceptible (R0.12) avocado rootstock infected with *P. cinnamomi* using an RNA-sequencing approach. The results showed that the partially resistant rootstock has increased expression levels of 84 *PaNLRs* observed up to 24 hours post-inoculation, while the susceptible rootstock only showed increased *PaNLR* expression during the first 6 hours post-inoculation. Results of this study indicate that the partially resistant avocado rootstock has a stronger, more prolonged ETI response which enables it to suppress *P. cinnamomi* growth and combat disease caused by this pathogen. Furthermore, the identification of *PaNLRs* may be used to develop resistant rootstock selection tools which can be employed in the avocado industry to accelerate rootstock screening programmes.

Introduction

Avocados (*Persea americana* Mill.) are an agriculturally important crop in many countries, including South Africa, Spain and Mexico (Bulagi *et al.* 2016; Vargas-Canales *et al.* 2020). The annual gross production value of avocados in South Africa increased by 14.2% in 2018-2019 to a total of R1.42 billion, when compared to 2017-2018. Phytophthora root rot, caused by the hemibiotrophic oomycete, *Phytophthora cinnamomi* Rands, remains the largest threat to the avocado industry, in countries where the pathogen is present (Hardham and Blackman 2018). The pathogen infects the fine feeder roots of avocado trees, leading to decreased water and nutrient transportation between cells (Coffey 1987). A decline in tree health is observed which ultimately leads to plant death. *P. cinnamomi* can survive in soils over long periods of time through the production of chlamydospores and oospores, thus limiting the number of effective control methods for Phytophthora root rot (Belisle *et al.* 2019; Dobrowolski *et al.* 2008). Phosphite trunk injections, use of partially resistant rootstocks and organic mulching practices are methods currently employed by the avocado industry to control *P. cinnamomi* (Giblin *et al.* 2005). However, research has shown that *P. cinnamomi* has the potential to become less sensitive towards phosphite trunk injections (Dobrowolski *et al.* 2008). Continued screening for *P. cinnamomi* resistant rootstocks is thus of utmost importance and can be accelerated when host-pathogen interactions are understood.

Plant immune responses influence host-pathogen interactions and involve a myriad of proteins which activate complex, multilayered signalling pathways in response to pathogen attack (Dangl and Jones 2001; Naveed *et al.* 2020). These can be categorized into two main responses; the PAMP (Pathogen associated molecular pattern) triggered immune response (PTI) and the Effector triggered immune response (ETI) (Davis and Hahlbrock 1987; Jones and Dangl 2006). The recognition of PAMPs by membrane-bound Pattern recognition receptors (PRRs) activate an innate immune response, which is lower in amplitude when compared to the ETI response, and forms part of the plant's first line of defence against pathogens (Matzinger 2007). Pathogens, in turn, produce effector proteins which are secreted into plant cells to interfere with this process. These effector molecules may then be recognized by intracellular proteins, such as Resistance (R) proteins, either directly or indirectly (Monteiro and Nishimura 2018). Upon effector recognition, R proteins are activated and trigger ETI– a high amplitude, robust immune response. The primary mode of action of ETI is to activate localized cell death caused by the Hypersensitive response (HR), aimed at arresting pathogen growth (Cui *et al.* 2015).

R proteins are classified into five diverse groups based on protein structure and domains (Bezerra-Neto *et al.* 2020). The largest group consists of proteins with Nucleotide binding and Leucine rich repeat domains, referred to as NLRs (McDowell and Woffenden 2003). Other groups include Receptor-like proteins (RLPs), Receptor-like kinases (RLKs) and Transmembrane Coiled-coil

proteins (TM-CCs). The NLR group can be further sub-divided into two classes, based on the NLR's N-terminus domain. The first class has a Coiled-coil (CC) domain, while the second class has a Toll/interleukin-1 receptor (TIR) structure domain. These NLRs are termed CNLs and TNLs, respectively. The CNL class also includes NLR proteins with both a CC domain and a RPW8 domain (resistance to powdery mildew), termed CC_R-NLRs or C_RNLs (Zong and Cheng 2016). *CNLs* are more abundant in the genomes of tree species when compared to *TNLs*, although the opposite is seen in *Arabidopsis* (Neale *et al.* 2017). Certain angiosperm and conifer species also do not follow this pattern due to *TNL* duplications, resulting in increased *TNL*:*CNL* ratios. Tandem duplications and *NLR* gene family expansions may have increased fitness levels of tree species that need long-term defence strategies against pathogens (Tobias and Guest 2014). As a result, these duplicated gene sequences are mostly found in gene clusters within plant genomes (Meyers *et al.* 2003). Head-to-head *NLR* genes may express proteins which interact to form homo– or heterodimers, often vital for proper NLR function (Liang *et al.* 2019). These NLR protein dimers greatly increase the pathogen recognition potential of different NLR protein complexes (Van Wersch and Li 2019).

A few NLR proteins are constantly expressed at low basal levels which allow plants to 'scan' for invading pathogens (Meyers *et al.* 2002). Most of the genes coding for NLR proteins, however, show differential expression patterns after pathogen attack. This is influenced by the species of pathogen, excreted effector proteins and the plant's genotype (Andam *et al.* 2020; Christie *et al.* 2016). In *Eucalyptus grandis* challenged by *Leptocybe invasa* and *Chrysoporthe austroafricana*, 218 and 343 *NLR*s were differentially expressed, respectively (Christie *et al.* 2016). RGA1, a TNL protein in the tree species *Salix viminalis*, showed higher expression in the resistant host when compared to its susceptible counterpart after *Melamspora larici-epitea* infection (Martin *et al.* 2016). Higher RGA1 expression allows for earlier ETI activation which ultimately leads to enhanced disease resistance. The level and timing of *NLR* expression is crucial, as this ultimately governs whether a plant would be successful in countering pathogen attack (Umadevi and Anandaraj 2017). Transgenic plants with higher *NLR* gene expression demonstrated increased resistance to plant pathogens, even when these plants were transformed with non-native *NLR* genes. Expression of *ZmNB25*, a *NLR* first identified in maize, increased the resistance levels of *Arabidopsis* and rice towards *Pseudomonas syringae* pv. *tomato* DC3000 and *Bipolaris maydis*, respectively (Xu *et al.* 2018). Understanding how the expression of *NLR*s change during pathogen infection, and subsequently influence disease resistance, is vital to understanding complex plant-pathogen interactions.

To date, 49 complete putative *NLR* genes have been identified in avocado using microarray and RNA-sequencing analysis (Pérez-Torres *et al.* 2021; Van den Berg *et al.* 2018). In the study done by Pérez-Torres *et al.* (2021), Hass avocado stems were infected with *Fusarium kuroshium*, which causes *Fusarium* dieback disease in avocado. However, only four *NLR* genes were differentially expressed after *F. kuroshium* infection. Additionally, only a single avocado *NLR* gene has been

implicated in the defence against *P. cinnamomi* infection in an avocado rootstock (Van den Berg *et al.* 2018). This *NLR*, functionally annotated as *RPP13*-like *protein 4*, showed increased expression after *P. cinnamomi* infection in a partially resistant rootstock. The use of RNA-seq and microarray data to identify *NLR* genes is limited by the fact that these genes need to be expressed to enable detection and identification. Avocado *NLR* gene identification has further been hampered by the lack of a high-quality genome assembly. Three avocado genomes, the Mexican landrace cultivar (*Persea americana* var. *drymifolia*), the Hass fruiting cultivar (Rendón-Anaya *et al.* 2019) and the VC75 rootstock (Avocado Genome Consortium, Article in preparation) have only recently been sequenced, providing an opportunity to identify significantly more *NLR* genes within the avocado genome.

The discovery of *NLR* genes within the avocado genome could provide novel insight into the interactions of this plant with various pathogens. The current study set out to identify avocado *NLR* genes using the available genome sequences, and subsequently assess their expression during *P. cinnamomi* infection of partially resistant and susceptible rootstocks. We identified 161 putative *PaNLR* genes in the VC75 rootstock avocado genome, based on amino acid sequences characteristic of conserved NLR domains. Furthermore, we analyzed the expression of the candidate *PaNLR* genes in a partially resistant and susceptible rootstock following *P. cinnamomi* inoculation using an RNA-seq approach. We found significantly higher expression levels of 84 *PaNLR* genes in the partially resistant rootstock when compared to the susceptible rootstock after *P. cinnamomi* inoculation. This knowledge may benefit future rootstock screening programmes aimed at increasing resistance levels towards *P. cinnamomi*. The *PaNLR* gene sequences identified in this study serve as an invaluable resource which can be used to pinpoint proteins which play a role in defence responses against other avocado pathogens.

Materials and methods

1. Putative *PaNLR* gene identification

The *P. americana* VC75 genome was obtained from the Avocado Genome Consortium (Article in preparation). Gene and protein names assigned during genome annotation (Peame105C00g000000) were abbreviated to PC00g000000. The Hass fruiting cultivar (*P. americana* cv. Hass; GCA_008087245.1) and Mexican rootstock (*P. americana* var. *drymifolia*; GCA_008033785.1) genomes (Rendón-Anaya *et al.* 2019) were obtained from GenBank (NCBI Genbank). Putative *R* genes were identified and classified using the Resistance gene analog (RGA) prediction pipeline, RGAugury (Li *et al.* 2016; downloaded in September 2020). The avocado VC75 genome, as well as whole genome protein sequences from the VC75, Mexican and Hass genomes were used as input with default parameters. The pipeline identifies conserved RGA sequences and domains using five programmes: BLAST v. 2.10.1 (Camacho *et al.* 2009), nCoil v. 2.2 (Lupas *et al.* 1991), InterProScan v. 5.52-86.0 (Zdobnov and Apweiler 2001), Pfam_scan v. 1.6 (Finn *et al.* 2010)

and Phobius v. 1.01 (Käll *et al.* 2004). Putative NLR proteins were classified based on the identified domains, namely Nucleotide binding site (NB), Coiled-coil domain (CC), Coiled-coil with RPW8 domain (CC_R), Toll/interleukin-1 receptor (TIR), and Leucine rich repeat domain (LRR). Here, N, C, T and L represent NB, CC, TIR and LRR domains, respectively. Thus, a protein classified as CNL has a CC, NB and LRR domain, and a CN protein only has a CC and NB domain. RLKs, RLPs and TM-CC classifications were annotated if the protein sequences contained a transmembrane domain. After identification and classification, protein functional annotation was done by performing BLASTp analysis in the non-redundant NCBI database. Searches were performed using an expected threshold value of 0.00001, with only the top hit for each candidate *NLR* gene being considered. If no significant match could be identified, proteins were annotated as Disease resistance-like (DRL) proteins.

2. *PaNLR* gene cluster identification

Gene clusters were defined based on appropriate definitions from Meyers *et al.* (2003), Kohler *et al.* (2008) and Christie *et al.* (2016). A gene cluster was defined as: a genomic region which contained three or more *NLR* genes, with less than nine other genes between adjacent *NLR* genes, and with two adjacent *NLR* genes being less than 250kb apart. The VC75 genome General feature format (GFF) file was used to indicate the distance and number of neighboring genes between *NLR* genes. The positions of *NLR* genes were visualized using CViT v. 1.3 (Cannon and Cannon 2011).

3. Phylogenetic analysis

Phylogenetic analysis was used to assess whether *NLR* genes from the same gene cluster has high sequence similarity. Phylogenetic analysis included 161 *P. americana* NB-domain protein sequences, 10 complete protein sequences from *Cinnamomum micranthum* f. *kanehirae* (RWR97694.1, RWR95032.1, RWR91786.1, RWR92004.1, RWR93015.1, RWR98067.1, RWR88343.1, RWR88103.1, RWR87020.1, RWR85657.1; Chaw *et al.* 2019) and one complete protein sequence from *Solanum bulbocastanum* (Q7XBQ9.1; Song *et al.* 2003). This *S. bulbocastanum* sequence was used since no RGA2 sequences were identified in *C. micranthum* f. *kanehirae* (Chaw *et al.* 2019). Sequence alignment was performed using ClustalW v. 2.1 with default parameters in MEGA X (Kumar *et al.* 2018; Thompson *et al.* 1994). A Maximum likelihood phylogenetic tree was produced using the Jones-Taylor-Thornton substitution model and 1000 bootstrap replications.

4. Plant inoculation and RNA-sequencing

NLR expression data was obtained by dual RNA-sequencing of *P. americana* inoculated with *P. cinnamomi*. Roots from partially resistant (Dusa®) and susceptible (R0.12) rootstocks were inoculated by dipping in *P. cinnamomi* (isolate GKB4) zoospore suspension with a concentration of

1.4 x 10^5 zoospores/ml. Thereafter, plants were replanted in a mixture of vermiculite and perlite (1:1 ratio) and roots were harvested at 6, 12, 24 and 120 hours post-inoculation (hpi). Three biological replicates from three independent plants were harvested at each time point. For control samples, three plants per rootstock were mock-inoculated using sterile water and root samples were harvested at 24 hpi.

Root samples were flash frozen using liquid N_2 and stored at -70°C. The samples were then powdered using an IKA® Tube Mill (IKA®, Staufen, DUE). Modified CTAB extractions were performed to extract total RNA (Chang *et al.* 1993). RNA extractions were purified using a Qiagen RNeasy clean up kit (Qiagen, Valecia, CA, USA), followed by DNase I treatments (Fermentas Life Sciences, Hanover, USA). An Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) was used to measure RNA purity and quality. Samples were stored at -70°C before being sent to Novogene (Novogene Corporation Inc., Chula Vista, CA, USA) for paired-end (250-300bp insert cDNA library) sequencing using Illumina Hiseq 2500 with PE150 mode.

5. Expression analysis

Dual RNA-sequencing data was analyzed by the Avocado Research Programme and used during this study (Article in preparation). Briefly, RNA-seq reads were trimmed and low-quality bases were removed using Trimmomatic v. 0.39 (Bolger *et al.* 2014). FASTQC v. 0.11.9 was used to confirm read quality and the resultant reports were summarized using MultiQC (Ewels *et al.* 2016). RNA-seq reads were aligned to the *P. americana* VC75 genome using HISAT v. 2.0.6 (Kim *et al.* 2015). Gene level transcript abundance was quantified using featureCounts v. 2.0.1 (Liao *et al.* 2014) during initial expression screens within RNA-seq libraries across all time-points (6 hpi, 12 hpi, 24 hpi, 120 hpi), using the mock-inoculated or susceptible rootstock libraries as a reference. DESeq2 (Love *et al.* 2014) was used for the normalization and analysis of counts. Quantification data for *NLR* genes were extracted using R studio v. 1.4.1106 (RStudio Team 2020) and gene IDs previously identified by RGAugury. Expression level differences were analyzed using two approaches: 1) comparing the expression of candidate *PaNLR* genes 6, 12, 24 and 120 hpi in both the susceptible and partially resistant rootstocks to that of their respective mock-inoculated samples, and 2) comparing the expression of candidate *PaNLR*s in the partially resistant rootstock to the expression in the susceptible avocado rootstock (mock-inoculated, 6, 12, 24 and 120 hpi). *PaNLR* genes were considered to be up– or downregulated when the Log_2 Fold Change (Log_2FC) value for each gene was ≥ 1 or ≤ -1, respectively. False discovery rate adjusted *p*-values ≤ 0.05, generated as part of the DESeq2 package, were used to indicate statistical significance. Heatmaps and dendrograms depicting expression level differences (Log_2FC) were generated using the Pheatmap package v. 1.0.12 (https://CRAN.R-project.org/package=pheatmap) in R studio v. 1.4.1106 (RStudio Team 2020). To assess whether *NLR* genes within a gene cluster were co-expressed, *NLR* expression data was analyzed using Clust v. 1.12.0 (Abu-Jamous and Kelly 2018)

Results

1. Putative *PaNLR* genes identified in the avocado genome

The RGAugury pipeline identified 259 putative *PaNLR* genes within the VC75 rootstock genome (Figure 1), while no *NLR* genes could be identified within the Mexican and Hass genomes. *NLR* gene sequences which did not include a LRR domain sequence were removed from further analysis and considered as incomplete *NLR* genes. This resulted in 161 *NLR* sequences which were classified as complete *NLR* genes. Of these genes, 102 were classified as *CNLs*, two as *C_RNLs*, 56 as *NLs* and one as a *TNL*, based on the domains present within their predicted amino acid sequences.

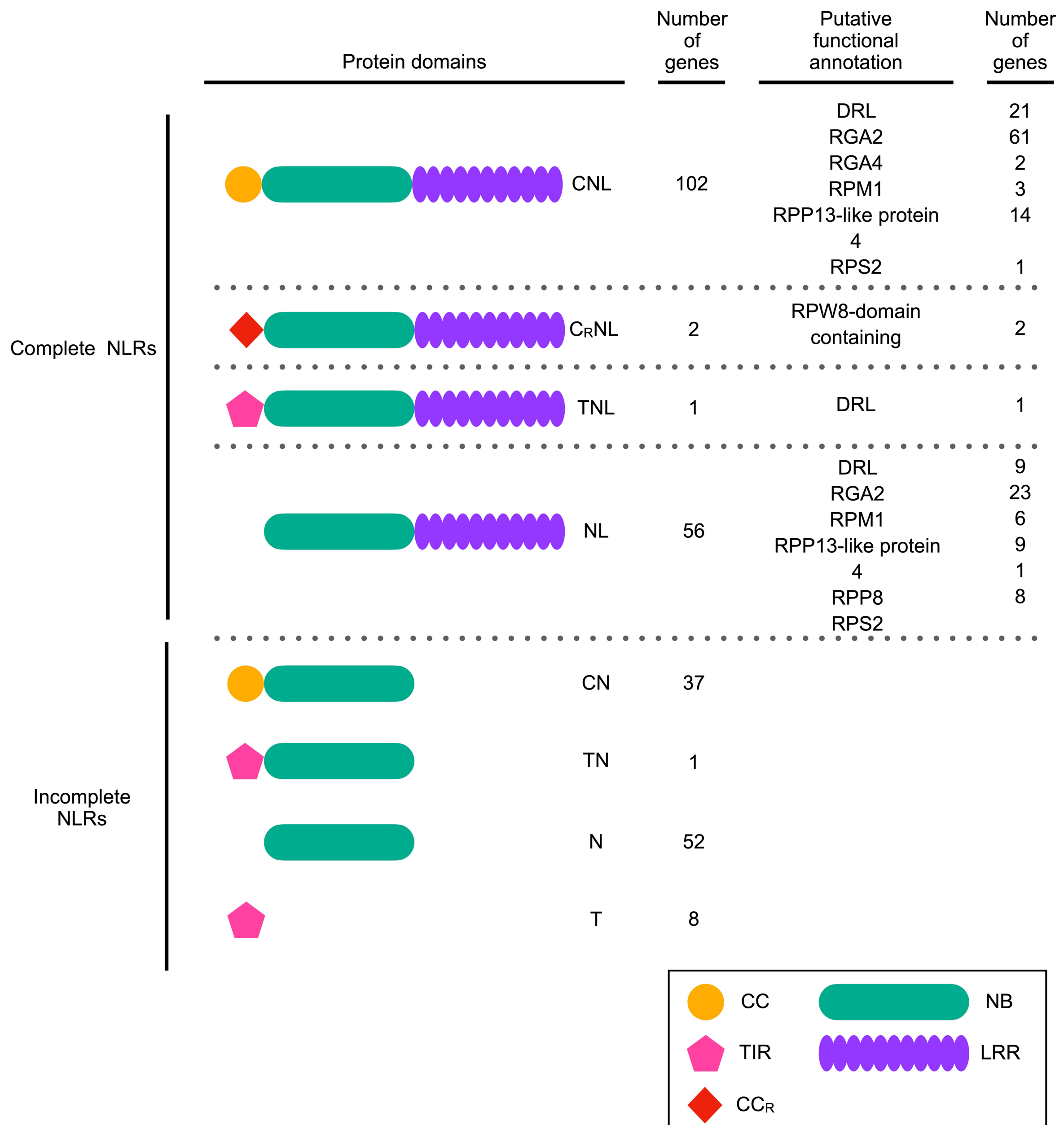


Figure 1. The number of *NLR* genes identified in the VC75 *Persea americana* genome and the set of protein domains each gene encodes for. Putative NLR protein functional annotations predicted using BLASTp analysis are also listed (C/CC – Coiled-coil domain; C_R/CC_R – Coiled-coil RPW8 domain; DRL – Disease resistance-like protein; L/LRR – Leucine rich repeat domain; N/NB – Nucleotide binding domain; T/TIR – Toll/interleukin-1 receptor domain).

Putative protein functional annotation of the 161 complete *PaNLR* gene candidates were assigned using BLASTp. In total, 31 sequences were assigned as DRL proteins. More than 52% of sequences were putatively identified as RGA2 proteins (Figure 1). Other sequence identifications included RGA4, RPM1, RPP13-like protein 4, RPP8, RPS2 and RPW8-domain-containing type proteins. The RGAugury pipeline also identified RLP, RLK and TM-CC proteins using VC75 whole genome protein sequences. These protein sequences were separated from NLR sequences if a transmembrane domain sequence was identified. In total, 106 RLP sequences, 889 RLK sequences and 189 TM-CC sequences were identified.

2. *PaNLR* gene clusters identified in the VC75 genome

PaNLR gene clusters were identified based on neighboring *PaNLR* genes being less than 250kb apart and having less than three non-*NLR* genes between them. In total, 13 *PaNLR* gene clusters were identified, accounting for 74 (45.9%) of the complete *PaNLR* gene sequences (Figure 2). Thirteen *PaNLR* genes were mapped to unanchored chromosomes and were thus excluded from the cluster analysis. Chromosome 2 had four gene clusters (the largest set of clusters on any of the chromosomes) and also contained the largest gene cluster (consisting of nine *PaNLR* sequences). No gene clusters were identified on chromosomes 4, 5, 8, 9, 10 and 12. Eight of the gene clusters contained sequences which encode RGA2 proteins, with the gene clusters occurring on chromosomes 6 and 7 lacking *PaNLR* genes encoding RGA2 proteins (Table 1).

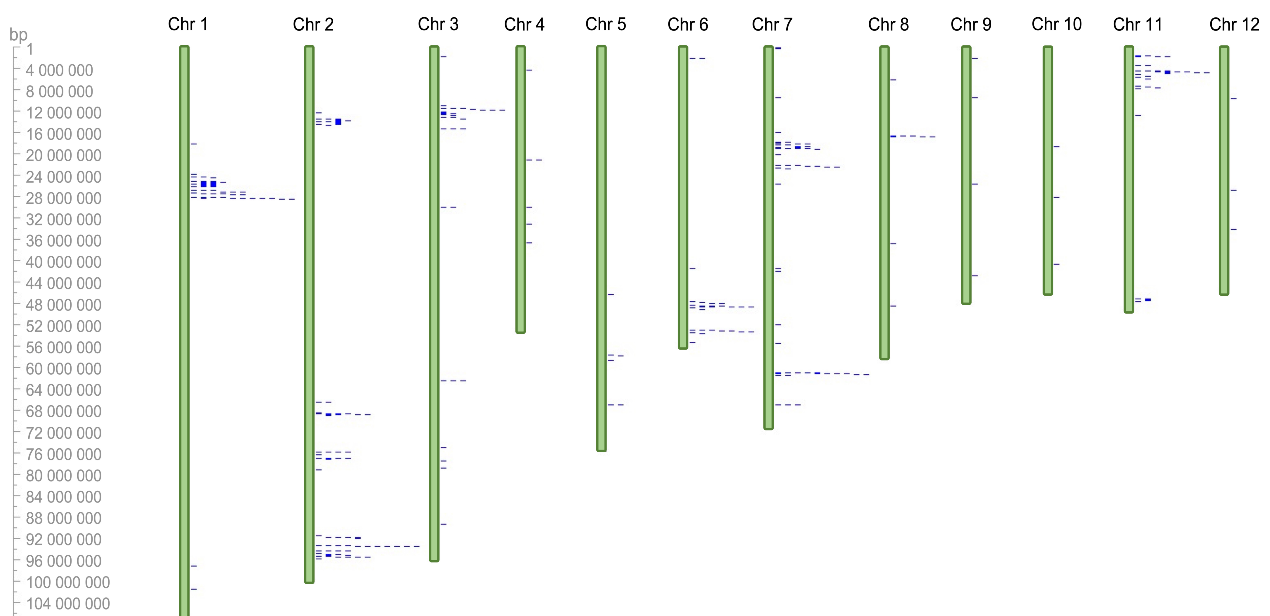


Figure 2. Chromosomal location of 148 putative *PaNLR* genes identified within the *Persea americana* VC75 genome (represented by blue marks). The genes were mapped to 12 chromosomes (green bars) using CViT. Chromosome 0 was excluded from the analysis as it is not representative of a true chromosome, thus 13 *PaNLR* genes could not be mapped to chromosomes 1-12 and are not shown in the figure.

Table 1. Types of resistance genes found within *PaNLR* gene clusters on different chromosomes within the genome of *Persea americana* (VC75).

Chromosome	Cluster	Number of *PaNLR* genes	Type of PaNLR proteins encoded
1	1	7	RGA2
	2	8	RGA2
2	1	5	RGA2
	2	3	RGA2
	3	6	DRL, RGA2 and RGA4
	4	9	RGA2
3	1	3	RGA2
6	1	6	DRL
	2	5	RPS2, DRL and RPP13
7	1	4	RPM1 and DRL
	2	8	RPP13
11	1	3	RGA2
11	2	7	RGA2

3. High *PaNLR* sequences similarity within *NLR* gene clusters

Phylogenetic analysis was performed to infer evolutionary relatedness between the 161 identified putative *PaNLR* genes, using NB-domain protein sequences (Figure 3). Complete NLR protein sequences from *C. micranthum* f. *kanehirae* and *S. bulbocastanum* were included in the analysis. The analysis revealed that *PaNLR* genes from the same *NLR* gene cluster grouped together within a clade, indicating high sequence similarity within clusters and possible gene duplication events. Most PaNLRs did not form a clade with *C. micranthum* f. *kanehirae* NLRs, indicating high diversification of *P. americana* NLRs after these two species diverged, especially RGA2 type PaNLRs.

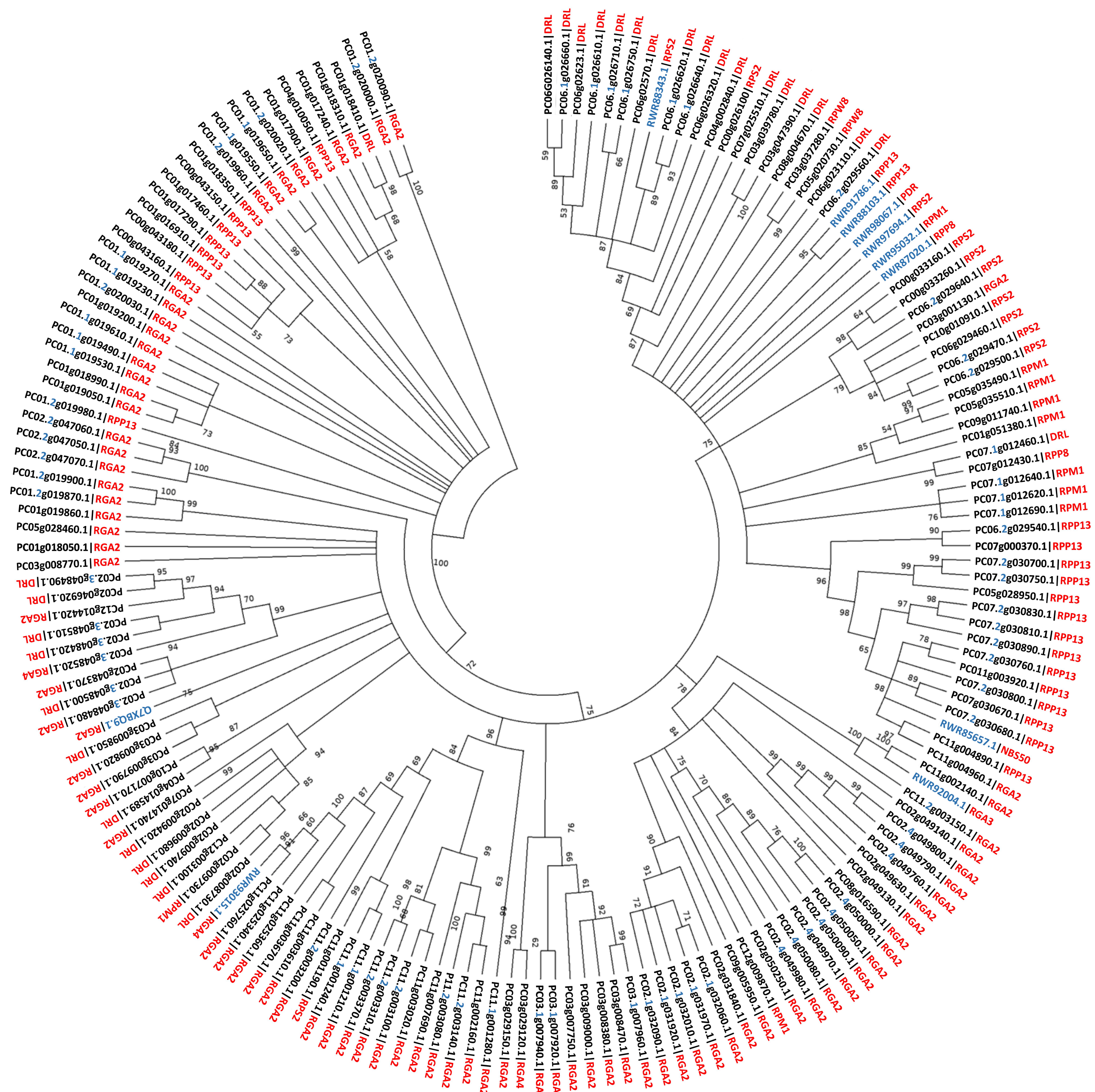


Figure 3. Phylogenetic relationship of 161 *Persea americana* (VC75) Nucleotide binding domains from putative *PaNLR* genes. Evolutionary history was inferred using the Maximum likelihood method and JTT matrix-based model following ClustalW alignment. A total of 1000 bootstrap replicates were performed, with bootstrap values over 50 being shown above branch points. NB-domain protein sequences of *P. americana* (PC) with complete NLR sequences from *Cinnamomum micranthum* f. *kanehirae* (RWR) and *Solanum bulbocastanum* (Q) were used during the analysis. *P. americana* identification numbers include the gene cluster number, where appropriate (in blue) and protein type (in red). Unidentified PaNLR protein types were termed Disease resistance-like (DRL) proteins. Sequences from other species also include protein type (NBS50 – NBS-LRR disease resistance protein NBS50; PDR – Disease resistance-like protein isoform X1).

4. *PaNLR* expression following *P. cinnamomi* inoculation

Expression analysis was performed using dual RNA-sequencing data obtained from partially resistant and susceptible avocado rootstocks inoculated with *P. cinnamomi* zoospores. In total, 145 of the 161 identified complete *PaNLR* genes were expressed in the roots of both rootstocks, across all timepoints. A clear difference in *PaNLR* expression was observed between the two rootstocks in response to *P. cinnamomi* inoculation. In the partially resistant rootstocks (Dusa®), a total of 84 *PaNLR* genes showed a significant ($p \leq 0.05$) change in expression level during at least one timepoint after *P. cinnamomi* inoculation, when compared to mock-inoculated samples (Figure 4). However, only 74 *PaNLRs* showed a significant ($p \leq 0.05$) change in expression in the susceptible rootstocks (R0.12) after inoculation, when compared to mock-inoculated samples. The number of *PaNLR* genes with expression level differences in response to *P. cinnamomi* inoculation differed most notably between the two rootstocks at 12 and 24 hpi (Table 2). Only six *PaNLR* genes were differentially expressed in R0.12 at 12 and 24 hpi, compared to 74 *PaNLR* in Dusa®. *PaNLR* genes within a cluster were shown not to be co-expressed based on the Clust analysis.

PC03g007960|RGA2 was the most upregulated *PaNLR* gene in Dusa®, with a Log_2FC value of 8.02 ($p < 0.01$) at 12 hpi (Figure 4). This gene was also upregulated in Dusa® at both 6 ($Log_2FC = 7.2$; $p < 0.01$) and 24 hpi ($Log_2FC = 7.8$; $p < 0.01$), while only being upregulated in R0.12 at 6 hpi ($Log_2FC = 7.8$; $p < 0.01$). PC03g009000|RGA2 was the most upregulated *PaNLR* gene in R0.12 at 6 hpi with the largest Log_2FC value of 8.46 ($p < 0.01$) of all samples. This gene did not show any significant changes in expression in any of the samples collected from Dusa®. Furthermore, PC11g001210|RGA2 and PC11g001240|RGA2 were upregulated in R0.12 at 24 hpi ($Log_2FC = 7.6$; $p < 0.01$ and $Log_2FC = 5.1$; $p < 0.05$, respectively), but did not show any significant change in expression in Dusa®, at any time point.

Table 2. Number of *PaNLR* genes expressed in two avocado rootstocks in response to *Phytophthora cinnamomi* inoculation at different timepoints post-inoculation, when compared to mock-inoculated rootstocks (hpi – hours post-inoculation).

	Partially resistant rootstock (Dusa®)		Susceptible rootstock (R0.12)		
Time (hpi)	Upregulated genes	Downregulated genes	Upregulated genes	Downregulated genes	Common between rootstocks
6	63	1	64	1	54
12	64	2	1	1	1
24	55	2	4	0	2
120	7	12	2	11	2

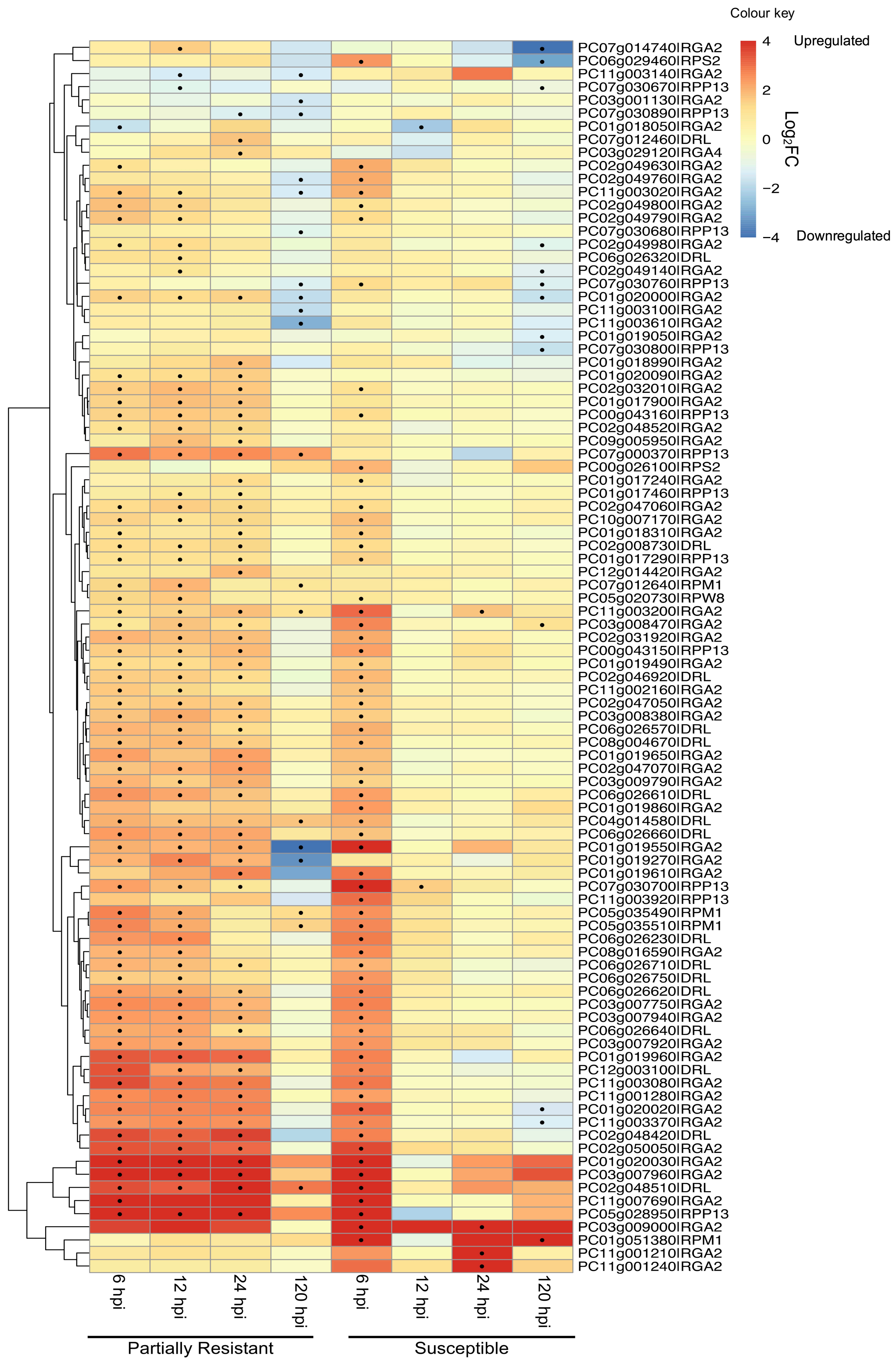


Figure 4. Heatmap and dendrogram showing the expression (as Log_2 Fold Change) of 94 *PaNLR* genes following *Phytophthora cinnamomi* inoculation of a partially resistant (Dusa®) and susceptible (R0.12) avocado rootstock. Dots indicate a significant change ($p \leq 0.05$ and $|Log_2FC| \geq 1$) in expression level when compared to mock-inoculated samples (hpi – hours post-inoculation).

When *PaNLR* gene expression was compared between the two rootstocks, with susceptible rootstock (R0.12) set as the reference, results indicated that *PaNLR* gene expression was higher in the partially resistant rootstocks (Dusa®), overall (Table 3). This was evident at the 12 and 24 hpi time points especially, with up to 74 *PaNLR* genes having higher expression ($p \leq 0.05$) in Dusa® at 12 hpi (Figure 5). PC11g001210|RGA2 and PC11g001240|RGA2 were two *PaNLR* genes that were expressed at significantly higher levels in Dusa® when compared to R0.12, in all samples collected including mock-inoculated roots, even though both *PaNLRs* were significantly upregulated at 24 hpi in R0.12 when compared to mock-inoculated samples (Figure 4). The Log_2FC values for both PC11g001210|RGA2 and PC11g001240|RGA2 were larger than 8.5 ($p < 0.01$) in all samples except at 24 hpi, where the Log_2FC values decreased to 3.8 and 6.2 ($p < 0.01$), respectively (Figure 5). The *PaNLR* gene with the highest expression level in R0.12 when compared with Dusa®, was PC02g009680|DRL. However, this *PaNLR* was only expressed at significantly higher levels in R0.12 at 6 hpi ($Log_2FC = -7.6$; $p < 0.01$), with no significant difference in expression levels being observed at any other time point.

Table 3. Number of *PaNLR* genes with significantly higher expression in either the partially resistant (Dusa®) or susceptible (R0.12) avocado rootstock before and following *Phytophthora cinnamomi* inoculation (hpi – hours post-inoculation).

Time (hpi)	Genes with higher expression in Dusa®	Genes with higher expression in R0.12
Mock-inoculated	10	9
6	9	7
12	74	3
24	61	7
120	16	11

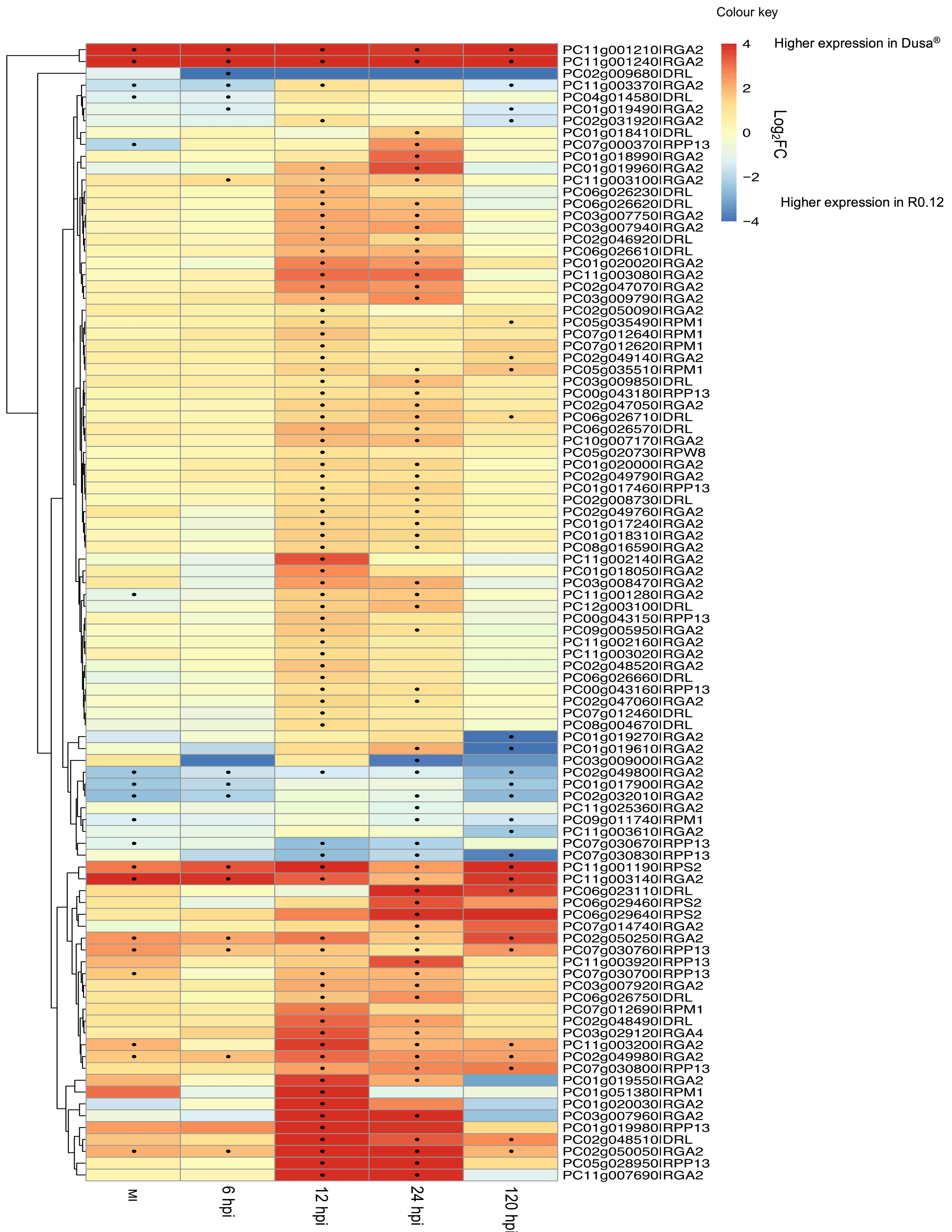


Figure 5. Heatmap and dendrogram showing *PaNLR* expression levels in a partially resistant avocado rootstock (Dusa®) mock-inoculated (MI) and following *Phytophthora cinnamomi* inoculation (hpi – hours post-inoculation) using a susceptible rootstock (R0.12) as the reference. A positive Log_2FC indicates higher expression in the partially resistant rootstock, while a negative Log_2FC indicates higher expression in the susceptible rootstock. Dots indicate a significant difference ($p \leq 0.05$ and $|Log_2FC| \geq 1$) in expression between the two rootstocks

Discussion

NLR proteins play a crucial role in plant immune responses by recognizing effector molecules produced by invading pathogens. Following effector recognition, NLR proteins activate ETI through complex signalling pathways, which leads to pathogen resistance (Monteiro and Nishimura 2018). NLR proteins have been studied extensively in many other crops including *S. bulbocastanum*, *Zea mays*, *Oryza sativa* and *Triticum monococcum* (Bozkurt *et al.* 2007; Collins *et al.* 1998; Lokossou *et al.* 2010; Mago *et al.* 1999). The results of these studies ultimately led to the breeding of crops with increased resistance towards various pathogens (Farnham and Baulcombe 2006; Wang *et al.* 2019). Thus far, a comprehensive set of *P. americana NLR* genes have not been identified, and moreover, *P. americana NLR* gene expression has never been studied during *P. cinnamomi* infection. Thus, a large knowledge gap remains in understanding ETI activation during *P. cinnamomi* infection in avocado rootstocks (Van den Berg *et al.* 2018). The knowledge of avocado NLR functionality is vital to understanding resistance towards *P. cinnamomi* in avocado rootstocks and can be used by the avocado industry for molecular breeding purposes.

Using the *P. americana* VC75 genome, we identified 161 putative complete *PaNLR* gene sequences. No *PaNLR* gene sequences could be identified in the Mexican and Hass genome assemblies, since these genomes are highly fragmented (Rendón-Anaya *et al.* 2019; Talavera *et al.* 2019). Of the 161 complete *PaNLR* sequences, 102 were classified as CNL proteins, two as C_RNL proteins, 56 as NL proteins and one as a TNL protein (Figure 1). The 56 gene sequences encoding NL proteins were found to be expressed during *P. cinnamomi* infection, indicating that these proteins may play a role in this host-pathogen interaction, even though they lack CC and TIR domains. *NLR* sequences lacking these motifs are also expressed in other plant species, including *E. grandis*, *Malus* x *domestica* and *Vitis vinifera*, further suggesting that these NLR proteins are still functional (Arya *et al.* 2014; Christie *et al.* 2016; Goyal *et al.* 2020). A higher *CNL*:*TNL* ratio was also observed in *E. grandis*, *M.* x *domestica* and *V. vinifera* woody species, making it unsurprising to observe a higher *CNL*:*TNL* ratio in *P. americana*. However, it was not expected that only one *TNL* sequence would be identified. This could be a result of the genome assembly and annotation programmes used not identifying full length gene sequences, thus producing truncated protein sequences as a result. BLASTp analysis was performed on the entire set of *P. americana* protein sequences and no additional *PaNLR* sequences could be identified. Furthermore, an independent study found only one *TNL* gene being expressed in Hass avocado stems during *F. kuroshium* infection (Pérez-Torres *et al.* 2021). This validated that no *TNL* motifs were missed during *PaNLR* identification using the RGAugury programme and VC75 genome.

Putative protein functional annotation revealed that more than 50% of the identified *PaNLR* genes encode RGA2-like proteins (Figure 1). This type of NLR protein was first identified in *S.*

bulbocastanum and is encoded for by *Rpi-blb1* (Van der Vossen *et al.* 2003). RGA2 proteins elicit an immune response and confer resistance towards *Phytophthora infestans* in potato and tomato plants, after recognizing ipiO RxLR proteins (Champouret *et al.* 2009). Recently, two *P. cinnamomi* RxLR proteins with high sequence similarity to *P. infestans* ipiO RxLRs were identified by Joubert *et al.* (2021). One of these RxLRs, PcinRxLR34a, was significantly upregulated in *P. cinnamomi* during infection of the susceptible rootstock R0.12, when compared to expression in mycelia. This suggests that this RxLR plays a role during pathogen infection. Future research should focus on identifying whether *P. americana* RGA2 proteins recognize these *P. cinnamomi* RxLRs.

RPP13-like protein 4 type proteins were the second largest group of PaNLR proteins identified in *P. americana*. RPP13-like protein 4 and RPP8 has been shown to confer resistance towards *Peronospora parasitica* and *Hyaloperonospora arabidopsidis*, respectively, in *Arabidopsis thaliana* (Bittner-Eddy *et al.* 1999; Mohr *et al.* 2010). *P. parasitica*, *H. arabidopsidis* and *P. cinnamomi* are oomycetes, suggesting that these pathogens may express Avirulence (Avr) proteins with similar structure and function (Cooke *et al.* 2000). This indicates that RPP13-like protein 4 and RPP8 in avocado may recognize *P. cinnamomi* effectors and play a role in rootstock resistance towards *P. cinnamomi*. The same assumption can be made regarding RPS2, which confers partial resistance towards *Phytophthora sojae*, a close relative to *P. cinnamomi*, in *Glycine max* (Mideros *et al.* 2007).

RPM1-like NLR proteins were also identified in avocado. Homologs of *RPM1*-like *NLR* genes in *A. thaliana* are responsible for recognizing *P. syringae* effectors during infection (Boyes *et al.* 1998). *P. syringae* has been isolated from avocados, however no symptoms of infection were observed (Scortichini *et al.* 2003). This might also explain why so few (5.6% of *NLRs*) *RPM1*-like genes were identified in avocado. Furthermore, since *P. syringae* infection does not present a threat to the avocado industry, *NLR* genes which confer resistance towards this pathogen would likely be of limited use in avocado screening programmes. Lastly, two PaNLRs were annotated as RGA4-like proteins; in *O. sativa*, RGA4 proteins form heterodimers with RGA5 proteins, which recognize *Magnaporthe oryzae* infection (Césari *et al.* 2013). RGA5 proteins acts as a receptor for *M. oryzae* Avr proteins and as a repressor of RGA4. Once RGA5 recognizes Avr proteins, RGA4 is released and activates cell death responses. Thus, in the absence of RGA5 proteins, RGA4 activates cell death in an Avr-independent manner (Césari *et al.* 2014). Since no *P. americana* proteins were identified as RGA5 proteins, it remains unclear whether the RGA4 proteins would respond to *P. cinnamomi* Avr proteins in avocado.

Gene cluster analysis was performed to identify possible duplication events of *P. americana NLRs* (Meyers *et al.* 2003). If *NLR* genes within a cluster were shown to be functionally important for rootstock resistance, *NLR* gene clusters can be targeted during molecular screening strategies. In total, 13 *PaNLR* gene clusters were identified in the *P. americana* genome (Table 1). Of these, four

clusters were identified on chromosome 2 with one containing nine *PaNLR* gene sequences. No clusters were observed on chromosomes 4, 5, 8, 9, 10 and 12 (Figure 1). Eight clusters only contained RGA2 protein sequences, indicating that these genes may have originated from gene duplication events as described by Meyers *et al.* (1998) and López *et al.* (2003). Retained *NLRs* following duplication indicate functional relevance, suggesting that these RGA2 NLRs may play an important role in avocado defence responses. In *Phaseolus vulgaris*, *RGA2* gene clusters were identified as Quantitative trait loci (QTL), which confer resistance towards *Colletotrichum lagenarium* (López *et al.* 2003). Further investigation focusing on functional significance will help identify whether the *PaNLR* gene clusters in *P. americana* can be used as QTL molecular markers during rootstock breeding programmes. Ultimately, these clusters serve as a reservoir for *NLR* diversity since duplicated genes are free to mutate, which may lead to novel NLRs being able to recognize novel effector proteins from pathogens (Innes *et al.* 2008).

Phylogenetic analysis revealed high similarity between *PaNLRs* within gene clusters, further indicating that *PaNLR* gene clusters may have originated from gene duplication events (Shao *et al.* 2014). During phylogenetic tree construction, 161 PaNLR Nucleotide binding domain protein sequences were used together with protein sequences from *C. micranthum* f. *kanehirae* and *S. bulbocastanum* (Chaw *et al.* 2019; Song *et al.* 2003). Sequences from *C. micranthum* f. *kanehirae* were used since this species is the closest relative to *P. americana* (both species form part of the Lauraceae family) in which NLRs have been identified (Wu *et al.* 2017). A RGA2 sequence from *S. bulbocastanum* was also included, since no RGA2 proteins were identified in *C. micranthum* f. *kanehirae* (Chaw *et al.* 2019). Phylogenetic analysis revealed that NB domain sequences within a *PaNLR* gene cluster grouped together, indicating high sequence similarity within these clusters (Figure 2). Moreover, few PaNLRs formed a clade with NLR sequences from *C. micranthum* f. *kanehirae*, indicating large *NLR* diversification within *P. americana* species. These observations might be the result of different pathogens shaping the *PaNLR* arsenal during the coevolutionary arms race between hosts and pathogens (Anderson *et al.* 2010).

Once putative *NLR* genes were identified in the VC75 genome, their expression was analyzed using dual transcriptomic data from partially resistant (Dusa®) and susceptible (R0.12) rootstocks inoculated with *P. cinnamomi*. Of the 161 *PaNLRs* identified in this study, 16 *PaNLRs* were not expressed in either rootstock at any timepoint. Many *NLRs* have tissue-specific expression levels in other plants, making these results unsurprising (Munch *et al.* 2018). Since this study investigated *PaNLR* expression in root tissues, it is expected that these 16 *PaNLRs* might play a role in recognizing pathogens which infect other avocado tissues. Interestingly, *PaNLR* genes within gene clusters did not show similar expression patterns after *P. cinnamomi* inoculation. This was also observed in *E. grandis* when infected with *C. austroafricana* and *L. invasa*. The authors attributed this to expressed *NLR* genes being functionally relevant, and not the result of being located within

active transcription zones by coincidence (Christie *et al.* 2016). Thus, we can hypothesize that *PaNLRs* in gene clusters being expressed following *P. cinnamomi* inoculation, do indeed have functional significance in activating defence responses against the invading pathogen.

During the first 6 hours of infection, more than 60 *PaNLR* genes showed a significant increase in expression in either rootstock, with a similar pattern of expression activation for 54 of the same *PaNLR* genes in both rootstocks (Figure 4). This indicates that both rootstocks have similar responses with regards to *PaNLR* expression during the first 6 hours of *P. cinnamomi* infection. *PaNLR* genes with the largest increase in expression at 6 hpi, were mainly RGA2 type proteins (PC01g020030, PC03g007960 and PC11g007690). RGA2 proteins activate the HR, and higher *RGA2* transcript levels were associated with increased *P. infestans* resistance in *S. bulbocastanum* (Bradeen *et al.* 2009). This upregulation of *RGA2* in both avocado rootstocks would likely result in a strong HR, which may limit *P. cinnamomi* growth.

PaNLR gene expression levels in Dusa® was higher when compared to R0.12, at both 12 and 24 hpi. Very few *PaNLR* genes showed differential expression patterns at 12 and 24 hpi in R0.12, which might indicate a decrease in ETI activation compared to Dusa®. Thus, the expression analysis revealed that Dusa® rootstocks overall have a stronger, more prolonged response to *P. cinnamomi* inoculation when compared to R0.12 rootstocks. *NLR* expression in susceptible varieties of *S. viminalis*, *C. arietinum* L. and *Brassica oleracea* do not show such stark differences in the expression when compared to resistant varieties, when infected with *Melampsora larici-epitea*, *Ascochyta rabiei* and *Fusarium oxysporum* f. sp. *conglutinans*, respectively (Andam *et al.* 2020; Liu *et al.* 2020; Martin *et al.* 2016). It was thus expected that a greater portion of *PaNLRs* would show increased expression in R0.12 at these timepoints. These results might be due to either the pathogen interfering with *PaNLR* expression, or the pathogen suppressing host responses in R0.12. For example, W boxes, which are *cis*-regulatory elements recognized by WRKY transcriptions factors, are often overrepresented in plant defence-related gene promoters including *NLR* promoter sequences (Mohr *et al.* 2010). In *A. thaliana*, *WRKY* expression was downregulated by Avr3a-type effectors from *Phytophthora parasitica* (Li *et al.* 2019). This would subsequently lead to decreased *NLR* expression. It would be interesting to see whether *P. cinnamomi* uses similar tactics to influence ETI in *P. americana.* Thus, investigating which *cis*-regulatory elements are shared between *PaNLR* genes would be of interest in future research. Moreover, *P. cinnamomi* RxLRs were shown to have increased expression levels at 12 and 24 hpi in R0.12 (Joubert *et al.* 2021). Since some RxLRs suppress programmed cell death, *P. cinnamomi* RxLRs could influence *PaNLR* expression and contribute to the results observed for R0.12 (Dalio *et al.* 2018). This data will help understand which PaNLR proteins might be important for recognizing *P. cinnamomi* effectors during infection and limiting *P. cinnamomi* growth. However, it must be noted that further studies, including protein-protein

interaction studies, are needed to concretely state which individual PaNLR proteins recognize *P. cinnamomi* effectors.

A previous study, also done on R0.12 and Dusa® rootstocks, showed that R0.12 had significantly higher *P. cinnamomi* pathogen loads when compared to Dusa®, at all tested time-points (Engelbrecht *et al.* 2013). The increased *PaNLR* expression in Dusa®, especially *RGA2 PaNLRs*, at 12 and 24 hpi is likely to increase the amplitude of ETI activation and the HR, assuming successful *P. cinnamomi* Avr detection. Studies have shown that overexpression of *NLR* genes leads to higher levels of resistance and subsequent decreased disease symptoms. In *N. benthamiana* plants, overexpression of the *Vitis amurensis NLR* gene, *VaRGA1*, resulted in increased resistance towards *P. parasitica* (Li *et al.* 2017). Two *RGA2 NLRs* (PC11g001210 and PC11g001240) showed much higher expression in Dusa® when compared to R0.12, in all samples (Figure 5). As described earlier, the RxLR in *P. cinnamomi* with high similarity to a RGA2 protein counterpart, ipiO1, showed increased expression in R0.12 at 12 hpi, when compared to mycelia control samples (Joubert *et al.* 2021). Since R0.12 RGA2 proteins are not upregulated at this timepoint, it may suggest that fewer of these RxLR effectors are recognized, resulting in a compromised HR. However, high expression of *RGA2 NLRs* in Dusa® at all timepoints could result in increased ETI and might lead to decreased pathogen growth rates and/or decreased zoospore germination. However, in R0.12, pathogen load could be higher due to decreased ETI. These differences might be why Dusa® is able to survive *P. cinnamomi* attack for longer periods of time and show less disease symptoms.

This study is the first to identify and classify putative *PaNLR* genes using the *P. americana* VC75 genome. Phylogenetic analysis revealed that many *PaNLRs* found within *NLR* gene clusters may have originated from gene duplication events. Up to 94 *PaNLR* genes showed expression differences in response to *P. cinnamomi* attack, indicating a possible role in *P. cinnamomi* recognition and ETI activation. Furthermore, *PaNLRs* showed sustained, increased expression in a partially resistant rootstock (Dusa®) after inoculation, which could explain how this rootstock is able to suppress *P. cinnamomi* growth. This research paves the way towards understanding *P. americana*-*P. cinnamomi* interactions on a molecular level. Future studies should focus on investigating protein-protein interactions between PaNLRs and *P. cinnamomi* Avr proteins, and how *P. cinnamomi* is able to suppress *PaNLR* expression in R0.12 rootstocks. Furthermore, future studies should also include functionally characterising the identified *PaNLRs* and investigating their role in defence responses against other *P. americana* pathogens.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

AF analyzed all data and drafted the manuscript. RB performed the early analysis of RNA-sequencing data and data curation. AB completed the assembly of the VC75 genome. JE designed and performed the experiments. VS and NvdB provided supervision of the study and revised the manuscript. All authors read and approved the final manuscript.

Funding

Project funding was generously provided by the Hans Merensky Foundation.

Acknowledgements

The authors appreciatively acknowledge the Avocado Genome Consortium for providing the *P. americana* VC75 rootstock genome. We would also like to thank Westfalia Fruit for plant material and the Hans Merensky Foundation for project funding.

Data Availability Statement

The datasets presented in this study can be found in online repositories. Consensus sequences for the candidate *P. americana NLR* genes identified in this study have been deposited in NCBI Genbank under accession numbers (pending). The RNA-seq data used in this study have been deposited in the Sequence Read Archive of NCBI Genbank under accession number PRJNA675400.

References

Abu-Jamous, B., and S. Kelly, 2018 Clust: automatic extraction of optimal co-expressed gene clusters from gene expression data. Genome Biology 19: 172.

Andam, A., A. Azizi, M. Majdi and J. Abdolahzadeh, 2020 Comparative expression profile of some putative resistance genes of chickpea genotypes in response to ascomycete fungus*, Ascochyta rabiei* (Pass.) Labr. Brazilian Journal of Botany 43: 123-130.

Anderson, J. P., C. A. Gleason, R. C. Foley, P. H. Thrall, J. B. Burdon *et al.*, 2010 Plants versus pathogens: an evolutionary arms race. Functional Plant Biology 37: 499-512.

Arya, P., G. Kumar, V. Acharya and A. K. Singh, 2014 Genome-wide identification and expression analysis of NBS-encoding genes in *Malus* x *domestica* and expansion of *NBS* genes family in Rosaceae. PLoS one 9: e107987-e107987.

Belisle, R. J., W. Hao, B. McKee, M. L. Arpaia, P. Manosalva *et al.*, 2019 New oomycota fungicides with activity against *Phytophthora cinnamomi* and their potential use for managing avocado root rot in California. Plant Disease 103: 2024-2032.

Bezerra-Neto, J. P., F. C. Araújo, J. R. C. Ferreira-Neto, R. L. O. Silva, A. N. C. Borges *et al.*, 2020 *NBS-LRR* genes-plant health sentinels: structure, roles, evolution and biotechnological applications, pp. 63-120 in *Applied Plant Biotechnology for Improving Resistance to Biotic Stress*, edited by P. Poltronieri and Y. Hong. Academic Press, Amsterdam.

Bittner-Eddy, P., C. Can, N. Gunn, M. Pinel, M. Tör *et al.*, 1999 Genetic and physical mapping of the *RPP13* locus, in *Arabidopsis*, responsible for specific recognition of several *Peronospora parasitica* (downy mildew) isolates. Molecular Plant-Microbe Interactions 12: 792-802.

Bolger, A. M., M. Lohse and B. Usadel, 2014 Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Boyes, D. C., J. Nam and J. L. Dangl, 1998 The *Arabidopsis thaliana RPM1* disease resistance gene product is a peripheral plasma membrane protein that is degraded coincident with the Hypersensitive response. Proceedings of the National Academy of Sciences of the United States of America 95: 15849-15854.

Bozkurt, O., E. E. Hakki and M. S. Akkaya, 2007 Isolation and sequence analysis of wheat *NBS-LRR* type disease resistance gene analogs using degenerate PCR primers. Biochemical Genetics 45: 469-486.

Bradeen, J. M., M. Iorizzo, D. S. Mollov, J. Raasch, L. C. Kramer *et al.*, 2009 Higher copy numbers of the potato *RB* transgene correspond to enhanced transcript and late blight resistance levels. Molecular Plant-Microbe Interactions 22: 437-446.

Bulagi, M., J. Hlongwane and A. Belete, 2016 Analysing the linkage between agricultural exports and agriculture's share of gross domestic products in South Africa. Journal of Agricultural Studies 4: 142.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Cannon, E. K. S., and S. B. Cannon, 2011 Chromosome visualization tool: a whole genome viewer. International Journal of Plant Genomics 2011: 373875.

Césari, S., H. Kanzaki, T. Fujiwara, M. Bernoux, V. Chalvon *et al.*, 2014 The NB-LRR proteins RGA4 and RGA5 interact functionally and physically to confer disease resistance. The EMBO Journal 33: 1941-1959.

Césari, S., G. Thilliez, C. Ribot, V. Chalvon, C. Michel *et al.*, 2013 The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors Avr-Pia and Avr1-CO39 by direct binding. The Plant Cell 25: 1463-1481.

Champouret, N., K. Bouwmeester, H. Rietman, T. Van der Lee, C. Maliepaard *et al.*, 2009 *Phytophthora infestans* isolates lacking class I ipiO variants are virulent on Rpi-blb1 potato. Molecular Plant-Microbe Interactions 22: 1535-1545.

Chang, S., J. Puryear and J. Cairney, 1993 A simple and efficient method for isolating RNA from pine trees. Plant Molecular Biology Reporter 11: 113-116.

Chaw, S. M., Y. C. Liu, Y. W. Wu, H. Y. Wang, C. Y. I. Lin *et al.*, 2019 Stout camphor tree genome fills gaps in understanding of flowering plant genome evolution. Nature Plants 5: 63-73.

Christie, N., P. A. Tobias, S. Naidoo and C. Külheim, 2016 The *Eucalyptus grandis NBS-LRR* gene family: physical clustering and expression hotspots. Frontiers in Plant Science 6: 1238.

Coffey, D. M., 1987 Phytophthora root rot of avocado– an integrated approach to control in California. California Avocado Society Yearbook 71: 121-137.

Collins, N. C., C. A. Webb, S. Seah, J. G. Ellis, S. H. Hulbert *et al.*, 1998 The isolation and mapping of disease resistance gene analogs in maize. Molecular Plant-Microbe Interactions 11: 968-978.

Cooke, D. E. L., A. Drenth, J. M. Duncan, G. Wagels and C. M. Brasier, 2000 A molecular phylogeny of *Phytophthora* and related oomycetes. Fungal Genetics and Biology 30: 17-32.

Cui, H., K. Tsuda and J. E. Parker, 2015 Effector triggered immunity: from pathogen perception to robust defence. Annual Review in Plant Biology 66: 487-511.

Dalio, R. J. D., H. J. Maximo, T. S. Oliveira, R. O. Dias, M. C. Breton *et al.*, 2018 *Phytophthora parasitica* effector PpRxLR2 suppresses *Nicotiana benthamiana* immunity. Molecular Plant-Microbe Interactions 31: 481-493.

Dangl, J. L., and J. D. G. Jones, 2001 Plant pathogens and integrated defence responses to infection. Nature 411: 826-833.

Davis, K. R., and K. Hahlbrock, 1987 Induction of defence responses in cultured parsley cells by plant cell wall fragments. Plant Physiology 84: 1286-1290.

Dobrowolski, M. P., B. L. Shearer, I. J. Colquhoun, P. A. O'Brien and G. E. S. Hardy, 2008 Selection for decreased sensitivity to phosphite in *Phytophthora cinnamomi* with prolonged use of fungicide. Plant Pathology 57: 928-936.

Engelbrecht, J., T. A. Duong and N. Van den Berg, 2013 Development of a nested quantitative real-time PCR for detecting *Phytophthora cinnamomi* in *Persea americana* rootstocks. Plant Disease 97: 1012-1017.

Ewels, P., M. Magnusson, S. Lundin and M. Käller, 2016 MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32: 3047-3048.

Farnham, G., and D. C. Baulcombe, 2006 Artificial evolution extends the spectrum of viruses that are targeted by a disease-resistance gene from potato. Proceedings of the National Academy of Sciences 103: 18828-18833.

Finn, R. D., J. Mistry, J. Tate, P. Coggill, A. Heger *et al.*, 2010 The Pfam protein families database. Nucleic Acids Research 38: D211-D222.

Giblin, F., K. Pegg, S. Willingham, J. Anderson, L. Coates *et al*., 2005 *Phytophthora* revisited, in Proceedings of the New Zealand and Australia Avocado Grower's Conference. Tauranga: New Zealand Avocado Growers Association.

Goyal, N., G. Bhatia, S. Sharma, N. Garewal, A. Upadhyay *et al.*, 2020 Genome-wide characterization revealed role of *NBS-LRR* genes during powdery mildew infection in *Vitis vinifera*. Genomics 112: 312-322.

Hardham, A. R., and L. M. Blackman, 2018 *Phytophthora cinnamomi*. Molecular Plant Pathology 19: 260-285.

Innes, R. W., C. Ameline-Torregrosa, T. Ashfield, E. Cannon, S. B. Cannon *et al.*, 2008 Differential accumulation of retroelements and diversification of *NB-LRR* disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiology 148: 1740-1759.

Jones, J. D. G., and J. L. Dangl, 2006 The plant immune system. Nature 444: 323-329.

Joubert, M., R. Backer, J. Engelbrecht and N. Van den Berg, 2021 Expression of several *Phytophthora cinnamomi* putative RxLRs provides evidence for virulence roles in avocado. PLoS ONE 16: e0254645.

Käll, L., A. Krogh and E. L. L. Sonnhammer, 2004 A combined transmembrane topology and signal peptide prediction method. Journal of Molecular Biology 338: 1027-1036.

Kim, D., B. Langmead and S. L. Salzberg, 2015 HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12: 357-360.

Kohler, A., C. Rinaldi, S. Duplessis, M. Baucher, D. Geelen *et al.*, 2008 Genome-wide identification of *NBS* resistance genes in *Populus trichocarpa*. Plant Molecular Biology 66: 619-636.

Kumar, S., G. Stecher, M. Li, C. Knyaz and K. Tamura, 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular Biology and Evolution 35: 1547-1549.

Li, P., X. Quan, G. Jia, J. Xiao, S. Cloutier *et al.*, 2016 RGAugury: a pipeline for genome-wide prediction of Resistance gene analogs (RGAs) in plants. BMC Genomics 17: 852.

Li, T., Q. Wang, R. Feng, L. Li, L. Ding *et al.*, 2019 Negative regulators of plant immunity derived from cinnamyl alcohol dehydrogenases are targeted by multiple *Phytophthora* Avr3a-like effectors. New Phytologist.

Li, X., Y. Zhang, L. Yin and J. Lu, 2017 Overexpression of pathogen-induced grapevine *TIR-NB-LRR* gene *VaRGA1* enhances disease resistance and drought and salt tolerance in *Nicotiana benthamiana*. Protoplasma 254: 957-969.

Liang, W., S. Van Wersch, M. Tong and X. Li, 2019 TIR-NB-LRR immune receptor SOC3 pairs with truncated TIR-NB protein CHS1 or TN2 to monitor the homeostasis of E3 ligase SAUL1. New Phytologist 221: 2054-2066.

Liao, Y., G. K. Smyth and W. Shi, 2014 featureCounts: an efficient general purpose programme for assigning sequence reads to genomic features. Bioinformatics 30: 923-930.

Liu, X., C. Zhao, L. Yang, M. Zhuang, Y. Zhang *et al.*, 2020 A time-resolved dual transcriptome analysis reveals the molecular regulating network underlying the compatible/incompatible interactions between cabbage (*Brassica oleracea*) and *Fusarium oxysporum* f. sp. *conglutinans*. Plant and Soil 448: 455-478.

Lokossou, A., H. Rietman, M. Wang, P. Krenek, H. Schoot *et al.*, 2010 Diversity, distribution, and evolution of *Solanum bulbocastanum* late blight resistance genes. Molecular Plant-Microbe Interactions 23: 1206-1216.

López, C. E., I. F. Acosta, C. Jara, F. Pedraza, E. Gaitán-Solís *et al.*, 2003 Identifying resistance gene analogs associated with resistances to different pathogens in common bean. Phytopathology 93: 88-95.

Love, M. I., W. Huber and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15: 550.

Lupas, A., M. Van Dyke and J. Stock, 1991 Predicting Coiled-coils from protein sequences. Science 252: 1162-1164.

Mago, R., S. Nair and M. Mohan, 1999 Resistance gene analogues from rice: cloning, sequencing and mapping. Theoretical and Applied Genetics 99: 50-57.

Martin, T., A. C. Rönnberg-Wästljung, J. Stenlid and B. Samils, 2016 Identification of a differentially expressed *TIR-NBS-LRR* gene in a major QTL associated to leaf rust resistance in *Salix*. PLoS ONE 11: e0168776.

Matzinger, P., 2007 Friendly and dangerous signals: is the tissue in control? Nature Immunology 8: 11-13.

McDowell, J. M., and B. J. Woffenden, 2003 Plant disease resistance genes: recent insights and potential applications. Trends in Biotechnology 21: 178-183.

Meyers, B. C., D. B. Chin, K. A. Shen, S. Sivaramakrishnan, D. O. Lavelle *et al.*, 1998 The major resistance gene cluster in lettuce is highly duplicated and spans several megabases. The Plant Cell 10: 1817-1832.

Meyers, B. C., A. Kozik, A. Griego, H. Kuang and R. W. Michelmore, 2003 Genome-wide analysis of NBS-LRR-encoding genes in *Arabidopsis*. The Plant Cell 15: 809-834.

Meyers, B. C., M. Morgante and R. W. Michelmore, 2002 TIR-X and TIR-NBS proteins: two new families related to disease resistance TIR-NBS-LRR proteins encoded in *Arabidopsis* and other plant genomes. Plant Journal 32: 77-92.

Mideros, S., M. Nita and A. E. Dorrance, 2007 Characterization of components of partial resistance, *RPS2*, and root resistance to *Phytophthora sojae* in soybean. Phytopathology 97: 655-662.

Mohr, T. J., N. D. Mammarella, T. Hoff, B. J. Woffenden, J. G. Jelesko *et al.*, 2010 The *Arabidopsis* downy mildew resistance gene *RPP8* is induced by pathogens and Salicylic acid and is regulated by W box *cis* elements. Molecular Plant-Microbe Interactions 23: 1303-1315.

Monteiro, F., and M. T. Nishimura, 2018 Structural, functional, and genomic diversity of plant NLR proteins: an evolved resource for rational engineering of plant immunity. Annual Review of Phytopathology 56: 243-267.

Munch, D., V. Gupta, A. Bachmann, W. Busch, S. Kelly *et al.*, 2018 The Brassicaceae family displays divergent, shoot-skewed *NLR* resistance gene expression. Plant Physiology 176: 1598-1609.

Naveed, Z. A., X. Wei, J. Chen, H. Mubeen and G. S. Ali, 2020 The PTI to ETI continuum in *Phytophthora*-plant interactions. Frontiers in Plant Science 11: 593905.

Neale, D. B., P. J. Martínez-García, A. R. De La Torre, S. Montanari and X. X. Wei, 2017 Novel insights into tree biology and genome evolution as revealed through genomics. Annual Review in Plant Biology 68: 457-483.

Pérez-Torres, C. A., E. Ibarra-Laclette, E. E. Hernández-Domínguez, B. Rodríguez-Haas, A. J. Pérez-Lira *et al.*, 2021 Molecular evidence of the avocado defence response to *Fusarium kuroshium* infection: a deep transcriptome analysis using RNA-seq. PeerJ 9: e11215.

Rendón-Anaya, M., E. Ibarra-Laclette, A. Méndez-Bravo, T. Lan, C. Zheng *et al.*, 2019 The avocado genome informs deep angiosperm phylogeny, highlights introgressive hybridization, and reveals pathogen-influenced gene space adaptation. Proceedings of the National Academy of Sciences 116: 17081.

RStudio Team, 2020 RStudio: Integrated Development for R. RStudio, PBC, Boston, MA.
Available online at: http://www.rstudio.com/

Scortichini, M., U. Marchesi, M. T. Dettori and M. P. Rossi, 2003 Genetic diversity, presence of the *syrB* gene, host preference and virulence of *Pseudomonas syringae* pv. *syringae* strains from woody and herbaceous host plants. Plant Pathology 52: 277-286.

Shao, Z. Q., Y. M. Zhang, Y. Y. Hang, J. Y. Xue, G. C. Zhou *et al.*, 2014 Long-term evolution of Nucleotide binding site-Leucine rich repeat genes: understanding gained from and beyond the legume family. Plant Physiology 166: 217-234.

Song, J., J. M. Bradeen, S. K. Naess, J. A. Raasch, S. M. Wielgus *et al.*, 2003 Gene *RB* cloned from *Solanum bulbocastanum* confers broad spectrum resistance to potato late blight. Proceedings of the National Academy of Sciences of the United States of America 100: 9128-9933.

Talavera, A., A. Soorni, A. Bombarely, A. J. Matas and J. I. Hormaza, 2019 Genome-wide SNP discovery and genomic characterization in avocado (*Persea americana* Mill.). Scientific Reports 9: 20137.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673-4680.

Tobias, P. A., and D. I. Guest, 2014 Tree immunity: growing old without antibodies. Trends in Plant Science 19: 367-370.

Umadevi, P., and M. Anandaraj, 2017 Genotype specific host resistance for *Phytophthora* in black pepper (*Piper nigrum* L.). Physiological and Molecular Plant Pathology 100: 237-241.

Van den Berg, N., W. Mahomed, N. A. Olivier, V. Swart and B. G. Crampton, 2018 Transcriptome analysis of an incompatible *Persea americana-Phytophthora cinnamomi* interaction reveals the involvement of SA– and JA-pathways in a successful defence response. PLoS ONE 13: e0205705.

Van Der Vossen, E., A. Sikkema, B. T. L. Hekkert, J. Gros, P. Stevens *et al.*, 2003 An ancient *R* gene from the wild potato species *Solanum bulbocastanum* confers broad-spectrum resistance to *Phytophthora infestans* in cultivated potato and tomato. The Plant Journal 36: 867-882.

Van Wersch, S., and X. Li, 2019 Stronger when together: clustering of plant *NLR* disease resistance genes. Trends in Plant Science 24: 688-699.

Vargas-Canales, J. M., G. Carbajal-Flores, T. I. Bustamante-Lara, J. H. Camacho-Vera, J. Fresnedo-Ramírez *et al.*, 2020 Impact of the market on the specialization and competitiveness of avocado production in Mexico. International Journal of Fruit Science 20: S1942-S1958.

Wang, L., L. Zhao, X. Zhang, Q. Zhang, Y. Jia *et al.*, 2019 Large-scale identification and functional analysis of *NLR* genes in blast resistance in the Tetep rice genome sequence. Proceedings of the National Academy of Sciences 116: 18479-18487.

Wu, C. C., F. H. Chu, C. K. Ho, C. H. Sung and S. H. Chang, 2017 Comparative analysis of the complete chloroplast genomic sequence and chemical components of *Cinnamomum micranthum* and *Cinnamomum kanehirae*. Holzforschung 71: 189-197.

Xu, Y., F. Liu, S. Zhu and X. Li, 2018 The maize *NBS-LRR* gene *ZmNBS25* enhances disease resistance in rice and *Arabidopsis*. Frontiers in Plant Science 9: 1033.

Zdobnov, E. M., and R. Apweiler, 2001 InterProScan– an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17: 847-848.

Zhong, Y., and Z. M. Cheng, 2016 A unique RPW8-encoding class of genes that originated in early land plants and evolved through domain fission, fusion, and duplication. Scientific Reports 6: 32923.