UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Sequencing ESTs of the avocado transcriptome to study the tolerant response to *Phytophthora cinnamomi*

by

Waheed Mahomed

Submitted in partial fulfilment of the requirements for the degree

Magister Scientiae

In the Faculty of Natural and Agricultural Sciences

Department of Genetics

University of Pretoria

Pretoria

Under the supervision of Dr. Noëlani van den Berg

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

© University of Pretoria

DECLARATION

I, the undersigned, hereby declare that the dissertation submitted herewith for the degree M.Sc. to the University of Pretoria, contains my own independent work and has not been submitted for any degree at any other university.

____

Waheed Mahomed

October 2011

© University of Pretoria

PREFACE

Avocado (*Persea americana* Mill.) is an important crop whose cultivation is severely threatened by *Phytophthora cinnamomi* Rands. The South African avocado industry makes an important contribution to the world's avocado supply, and is one of the world's largest exporters. If the current Phytophthora root rot problem is not addressed soon, the losses encountered by the avocado industry may become so extensive that it results in job losses. The scant information that is available for *P. cinnamomi* interaction studies indicate that there is no gene-for-gene interaction yet described between the pathogen and host. Avocado genomics are not well understood either and there is not much sequence data available for this basal angiosperm. The data available comprises of sequence that was generated in marker studies on fruit and flowering organs. It is now possible to generate large amounts of sequence data using high-throughput sequencing platforms and identify defence-related genes. The identification of defence-related genes in a tolerant rootstock will allow us to characterize the avocado-*P. cinnamomi* interaction on a molecular level.

The aim of this MSc was to identify defence-related genes in a tolerant rootstock and characterize their expression in order to understand the avocado-*P.* cinnamomi interaction.

Chapter 1 provides a comprehensive overview of the advances in molecular work conducted on avocado thus far. A background of avocado rootstock development is provided with details of molecular markers developed for use in avocado. Additionally, an introduction is also given to high-throughput sequencing and its application to non-model crops such as avocado. **Chapter 2** describes the mRNA isolation and EST pyrosequencing of avocado roots. Gene annotation of metabolic, cell wall associated and stress response genes are provided along with the characterisation of defence-related genes.

© University of Pretoria

Chapter 3 reports of the expression profiling of defence-related genes obtained from avocado root ESTs. The expression of nine defence-related genes are studied over six time points in *P. cinnamomi* infected R0.09 tolerant avocado roots. **Chapter 4** provides a general discussion of the result obtained in this study along with future applications of the sequencing data produced.

Preliminary results of this study were presented in the form of posters and oral presentations at various national and international meetings:

- Mahomed W, Myburg A. A, van den Berg N., (2011) Moving towards understanding defence in avocado-*Phytophthora cinnamomi* interactions. 47th congress of the Southern African Society for Plant Pathology (SASPP), January 23–26. Kruger National Park, South Africa.

- Mahomed W, Myburg A. A, van den Berg N., (2010) Sequencing the avocado transcriptome. 36th Annual Conference of the South African Association of Botanists (SAAB), January 11–14. North-West University, South Africa.

- Mahomed W, Myburg A. A, van den Berg N., (2010) Moving towards understanding defence in avocado-*Phytophthora cinnamomi* interactions. ACGT Regional Plant Biotechnology Forum Meeting, October 15. University of Pretoria, South Africa,

- Mahomed W, Myburg A. A, van den Berg N., (2011) Sequencing ESTs from *Phytophthora cinnamomi* infected avocado root cDNA libraries. VII World Avocado Congress September 4 – 9. Cairns, Australia.

© University of Pretoria

The findings presented in this M.Sc. dissertation represent the outcomes of a study undertaken from March 2009 to December 2010 in the Department of Genetics, University of Pretoria, under the supervision of Dr. Noëlani van den Berg. Chapters 2 and 3 have been accepted and published in **BMC Plant Biology (2011) 11, 167.**

© University of Pretoria

ACKNOWLEDGEMENTS

I would like to express my appreciation and gratitude to the following people and institutes for assisting me in the completion of this project:

- Dr. Noëlani van den Berg. Thank you for your insightful leadership and help throughout this degree, for your wisdom, kindness and unbelievable patience in assisting me in the finalization of this dissertation.

- My colleagues for their support and assistance in times of need.

- The National Research Foundation of South Africa (NRF), for funding.

- The Human Resources and Technology for Industry Programme (THRIP), for financial support of the research.

- The Hans Merensky Foundation for research funding.

- My family and friends that endured my hardships with me and gave hope when there was none.

© University of Pretoria

I would like to dedicate this thesis to my parents who have given me their unwavering support throughout my studies.

© University of Pretoria

Abbreviations and symbols:

%	Percentage
°C	Degree Celsius
ABA	Abscisic Acid
ABC	ATP-Binding Cassette
ADP	Adenosine Diphosphate
AMV	Avian Myeloblastosis Virus
APX	Ascorbate Peroxidase
ATP	Adenosine Triphosphate
bp	Base Pairs
cDNA	Complementary DNA
ChIP-seq	Chromatin Immunoprecipitation Sequencing
cm	Centimetre
COG	Clusters of Orthologous Groups
CT	Cycle Threshold
CTAB	Cetyltrimethylammoniumbromide
dCAS	Desktop cDNA Annotation System
DFP	DNA Fingerprinting
DNA	Deoxyribonucleic Acid
dNTPs	Deoxynucleotide Triphosphate

© University of Pretoria

DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic Acid
ESTs	Expressed Sequence Tags
F3H	Flavanone-3-Hydroxylase
g	Gram
g/L	Gram per Liter
G3PDH	Glyceraldehyde-3-Phosphate Dehydrogenase
Gb	Gigabases
GO	Gene Ontology
GRPs	Glycine-Rich Proteins
hpi	Hours Post Infection
hr	Hours
HR	Hypersensitive Response
HRGPs	Hydroxyproline-Rich Glycoproteins
HTS	High Throughput Sequencing
Kb	Kilobases
kDa	Kilodalton
kg	Kilogram
KOG	Eukaryotic Orthologous Groups
L	Liter
LRR	Leucine-Rich Repeats
Mb	Megabases

© University of Pretoria

µL	Microliter
MDP	Multidrug Resistance
min	Minute(s)
mM	Millimolar
mm	Millimeter
MPS	Massively parallel sequencing
mRNA	Messenger RNA
MRP	Multidrug Resistance-Associated Protein
MRP	Multidrug Resistance-Associated Protein
NBS	Nucleotide Binding Site
NCBI	National Center for Biotechnology Information
ng/ µl	Nanogram per Microliter
NGS	Next Generation Sequencing
NO	Nitric Oxide
NR	Non-Redundant
nt	Nucleotide
OBP	Oxysterol-Binding Protein
Pc	*Phytophthora cinnamomi*
PCD	Programmed Cell Death
PCR	Polymerised Chain Reaction
PDR	Pleiotropic Drug Resistance
PDR	Pleiotropic Drug Resistance

© University of Pretoria

PGN	Plant Genome Network
PRR	Phytophthora Root Rot
qRT-PCR	Quantitative Real-Time PCR
QTLs	Quantitative Trait Loci
RACE	Random Amplification of cDNA-ends
RAPDs	Random Amplified Polymorphic DNA Markers
RFLPs	Restriction Fragment Length Polymorphisms
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
rRNA	Ribosomal RNA
RT	Reverse Transcriptase
SA	Salicylic Acid
SAGE	Serial Analysis of Gene Expression
sec	Second(s)
SNP	Single Nucleotide Polymorphism
SSH	Selective Subtractive Hybridization
SSRs	Single Sequence Repeats
TAE	Tris-acetate EDTA
TBP	TATA-box Binding Protein
TCTP	Translationally Controlled Tumour Protein
TLP-PA	Thaumatin-Like Protein
U /µl	Units per Microliter

© University of Pretoria

USP	Universal Stress Protein
VNTRs	Variable Number of Tandem Repeats

© University of Pretoria

Table of Contents

Declaration i

Preface ii

Acknowledgements v

Abbreviations and symbols vii

Table of contents xii

Chapter 1: Molecular advances in avocado research 1

1.1 Introduction 2

1.2 Avocado 3

1.2.1 Avocado rootstock breeding 8

1.2.2 Avocado genomics 11

1.2.3 Molecular markers in avocado 12

1.2.3.1 Isozymes 13

1.2.3.2 RFLPs 14

1.2.3.3 RAPD 16

1.2.3.4 VNTRs 17

1.2.3.5 Avocado genetic linkage map 21

1.2.3.6 QTLs 22

© University of Pretoria

1.3 High-throughput sequencing platforms....25

1.3.1 SOLiD....27

1.3.2 Illumina....28

1.3.3 454 pyrosequencing....29

1.3.4 Applications....33

1.3.4.1 Transcriptome analysis....34

1.3.4.2 De novo sequencing....36

1.4 Conclusion....41

1.5 References....43

Chapter 2: EST sequencing of the avocado transcriptome in response to *Phytophthora cinnamomi*....54

2.1 Introduction....55

2.2 Materials & Methods....58

2.2.1 Avocado inoculation with *P. cinnamomi* & harvesting of root material....58

2.2.2 RNA extraction....58

2.2.3 mRNA isolation & cDNA synthesis....59

2.2.4 Pyrosequencing & data annotation....61

2.3 Results & Discussion....62

2.3.1 Avocado inoculation with *P. cinnamomi*....62

2.3.2 RNA extraction....63

© University of Pretoria

2.3.3 mRNA isolation & cDNA synthesis....63

2.3.4 Pyrosequencing & data annotation....63

2.3.4.1 EST identification & classifications....65

2.3.4.2 Transcripts involved in primary metabolism....67

2.3.4.3 Transcripts involved in stress related responses....71

2.3.4.4 Transcripts involved in the plant cell wall....75

2.3.5 Sequence alignments of selected defence-related ESTs....80

2.3.6 Species similarity between avocado and other plants....83

2.4 Conclusion....84

2.5 References....87

2.6 Figures & Tables....101

Chapter 3: Quantification of avocado defence-related genes in response to *Phytophthora cinnamomi*....150

3.1 Introduction....151

3.2 Materials & Methods....154

3.2.1 Avocado inoculation with *P. cinnamomi* and harvesting of root material....154

3.2.2 Generation of cDNA template....154

3.2.3 Quantitative gene expression analysis....155

3.2.4 Statistical analysis....156

3.3 Results....156

3.3.1 Symptoms of plants....156

© University of Pretoria

3.3.2 cDNA results.......157

3.3.3 Quantitative gene expression analysis.......157

3.4 Discussion.......159

3.5 References.......164

2.6 Figures & Tables.......171

Chapter 4: General discussion.......179

© University of Pretoria

CHAPTER 1

Molecular advances in avocado research

© University of Pretoria

1.1 Introduction

The avocado (*Persea americana* Mill.) is native to the western hemisphere and belongs to the *Lauraceae* family. The South African avocado industry makes an important contribution to the world's avocado supply, and is one of the world's largest exporters. The plant has a genome size of 8.83 x10^8 bp (bp) and the genome sequence is not publically available. One of the most important problems associated with avocado breeding is lack of rootstock tolerance to the soil-borne oomycete *Phytophthora cinnamomi* Rands, which causes Phytophthora root rot (PRR) (Coffey, 1987; Zentmyer, 1984; Zentmyer, 1955). Other problems with rootstock breeding involve the time associated with rootstock breeding and the lack of molecular knowledge available for the rootstock. In order to identify possible markers for rootstock improvement DNA sequence data is essential. Currently, there is not a large amount of data available for the identification of possible resistance traits and avocado sequence data needs to be generated to further the development of rootstocks.

The feasibility of using molecular markers for avocado improvement was established in 1989 (Clegg & Davis, 1989) with isozymes being used as genetic markers. Other markers such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPD) markers, single sequence repeats (SSRs), DNA fingerprinting (DFP) and VNTR (Variable Number of Tandem Repeats) were also developed to aid in fruit cultivar identification or to be used in the selection of desired traits (Acheampong *et al.*, 2008; Alcaraz & Hormaza, 2007; Clegg & Davis, 1989; Clegg *et al.*, 1992; Kobayashi *et al.*, 2000; Lavi *et al.*, 1994; Lavi *et al.*, 1991; Sharon *et al.*, 1997). The level of heterozygosity found in avocado makes it difficult to predict offspring performance (Davis *et al.*, 1998; Mhameed & Sharon, 1996) and despite the ongoing work in marker development, molecular work still needs improvement (Chen *et al.*,

© University of Pretoria

2007). In the study by Clegg *et al* (1999) it was shown that RFLP markers could possibly be used to select rootstocks with different tolerance levels to Phytophthora Root Rot (PRR).

Although there is much known about the cultivation of avocado not much is known about the tolerance of the plant to *P. cinnamomi.* A better understanding of these molecular level defence responses will result in the production of molecular tools for tolerant species identification, enhanced breeding programmes and molecular diagnostic tools. High-throughput sequencing has given molecular biology a new improved means of dissecting the genomes and transcriptomes of sequenced and uncharacterized organisms alike. It is now possible to characterise non-model organisms cheaply and effectively allowing a better understanding of changes in expression to abiotic or biotic stresses (Vera *et al*., 2008).

This literature review will provide background on avocado, its production and its available genetic resources along with the identification of high-throughput sequencing as a means to contribute to the molecular knowledge of this important basal angiosperm. The 454, Illumina and SOLiD platforms will be elaborated on with specific reference to transcriptome profiling, however this literature review will not encompass computational models and assembly tools.

1.2 Avocado

Avocado is native to Central America and Mexico with all the cultivated varieties being included in the species named *P. americana* and belonging to the genus- *Persea*, subgenus- *Persea*, family- *Lauraceae*. The *Lauraceae* family falls under the clade of magnolids that are sister to eudicot and monodicot clades. The family derives its name from the Grecian laurel tree, whose leaves were used to adorn the foreheads of competitors in the Olympic Games. *Persea americana* originated in the new world hence the name (Bergh & Ellstrand, 1986).

© University of Pretoria

There are three of eight varieties (geographical ecotypes) of *P. americana* that constitute commercial avocado production; the Mexican race, *P. americana* var. *drymifolia* (Schltdl. & Cham), the Guatemalan race, *P. americana* var. *guatemalensis* (L.O. Williams) and the West Indian race, *P. americana* var. americana (Ben-Ya'acov & Michelson, 1995; Chanderbali *et al.*, 2008). These are referred to as horticultural races. The three varieties are not different enough to constitute separate species but at the same time are too different to just be separate forms (Bergh & Ellstrand, 1986). Avocado diversity studies have been conducted and there is now genetic evidence in the form of DNA fingerprints to support the idea of three horticultural races (Lavi *et al.*, 1991). Research has been aimed at studying the basis of classification into these botanical races/ varieties.

The tree grows to 20 meters with greenish-yellow flowers 5–10 mm wide. The pear-shaped fruit is light green when unripe and shows variation between dark green and chocolate-brown when ripe, 7–20 cm long, weighs between 100 and 1000 g, is long-necked, sometimes round or oval and possesses a large central seed. The fruit is more or less round or pear-shaped having a fatty, flavourful, smooth, almost creamy textured flesh. Avocado is high in mono-unsaturated fat, has more potassium than bananas, and is rich in vitamins B, E and K., The fruit shows varying skin thickness between the Guatemalan and Mexican races with the Mexican race being the hardier of the two (Traub *et al.*, 1937). Characteristics such as oil content, blooming season and skin texture vary between the three avocado races (Table 1).

© University of Pretoria

Table 1. Characteristics of the three different avocado races.

Characteristics	West Indian	Guatemalan	Mexican
Origin	Tropical lowland	Tropical highland	Tropical highland
Blooming season	February to March	March to April	January to February
Fruit size	0.5 to 2 kg	0.2 to 2 kg	< 0.5 kg
Skin texture	Leathery-smooth	Woody-rough	Papery-smooth
Fruit oil content	Low	Medium to high	Medium to high
Fruit ripening	On-tree storage	On-tree storage	On-tree storage

Adapted from Bergh & Ellstrand, (1986) and Traub *et al.*, (1937)

Avocado production generates large revenue with the crop now grown commercially in more than 50 countries including Mexico, Indonesia, Colombia and the United States of America. The fruit only became a commercial crop after ‘Fuerte’ was selected from Mexico and introduced to California in 1911. The South African avocado industry makes an important contribution to the world’s avocado supply, and is one of the world’s largest exporters. In 2009 the total avocado production in South Africa was 75, 932 tonnes (http://faostat.fao.org).

Avocado production relies on a process referred to as grafting where the desired fruiting cultivar (scion) is attached to the top of a specific rootstock. An important key to grafting is the proper positioning of the scion on the rootstock (Elam, 1997). The benefit of grafting is to obtain the best rootstock cultivar for the soil and the fruiting scion with desired traits which ultimately results in higher production levels. Aside from rootstocks that may provide disease or pest resistance and scions with high fruit production, their interaction is reflected in the productivity of the avocados.

© University of Pretoria

The rootstock/scion compatibility has been studied profusely and it has been noted that the rootstock effect on tree size and productivity varies between different rootstocks, scions, and rootstock/scion interactions. Although there are no known compatibility issues between the three races of *P. americana*, some rootstocks/scions might be a much more vigorous grower and overpower its grafted counterpart or the rootstock might not be well adapted to a certain cultivar resulting in poor productivity of the combination. In the 1960s it was first demonstrated that production can be improved by using specific rootstock/scion combinations (Ben-Ya'acov & Michelson, 1995). The Duke 7 rootstock was indicated to be the most productive rootstock when grafted to the Hass scion (Whiley *et al.*, 1990). This illustrated the difference in productivity dependent on interactions between rootstocks and fruiting scions.

There are two types of rootstocks currently being used. Many countries such as Australia, Brazil, the Dominican Republic, Cuba and most other Caribbean countries as well as Florida in the United States use seedling rootstocks while South Africa, Mexico and Spain use clonal rootstocks (Ben-Ya'acov & Michelson, 1995). Clonal plants are produced by firstly grating the desired rootstock onto a germinated mother seed. The plantlet is placed in the dark and allowed to root after which the scion is grafted onto the rootstock. This is done due to the inherent difficulty in getting shoot cuttings to develop roots and also due to the non-uniformity of seeded rootstocks which has lead to clonal rootstock propagation (Ben-Ya'acov & Michelson, 1995). There are many advantages and disadvantages associated with each type of rootstock such as disease resistance, uniformity in flowering and production and sometimes seedling rootstocks are chosen due to availability and nursery performance but it has been noted that clonal rootstocks will become commercially widespread due to their advantages in a similar manner that clonal rootstocks revolutionised other fruit industries (Whiley *et al.*, 1990).

© University of Pretoria

There are many factors that affect avocado cultivation. Some of these are climatic factors, salinity stresses and aeration stresses that are soil dependent (Ben-Ya'acov & Michelson, 1995). The fruit crop is also susceptible to numerous diseases varying from viral, bacterial and fungal backgrounds. These diseases affect either the trunk or stem of the scion or the root system. Dothiorella rot (*Dothiorella gregaria/ Botrysphaeria ribis*) infects the fruit as well as the stem of the scion while avocado canker (*Xanthomonas campestris*) and girdling canker (*Cylindrodarpon* sp.) infects the stem of the scion. Common root diseases are Armillaria root rot (*Armillaria mellea*), Dematophora or Rosellinia root rot (*Dematophora necatrix*), the vascular Verticillium wilt (*Verticillium albo-atrum*) and Phytophthora root rot caused by *Phytophthora cinnamomi* (Coffey, 1987; Faber *et al.*, 2010; Zentmyer, 1984; Zentmyer, 1955).

Phytophthora cinnamomi is undoubtedly the most limiting disease of avocado production. It is a pathogenic oomycete belonging to the order *Perenosporales* (Linde, 1999). In 1922 R.D. Rands isolated *Phytophthora cinnamomi* from cinnamon trees in Sumatra and in 1942 Zentmyer concluded that this was the cause of the root rot and death of the trees (Coffey, 1987; Govers & Gijzin, 2006; Keane, 1981). The pathogen is controlled by using clean planting material and by using high organic matter soils (Giblin *et al.*, 2005). There are various efforts to combat the pathogen. A chemical means of controlling Phytophthora root rot is the use of Metalaxyl and by trunk injection of phosphorous acid into the plant, currently the use of phosphonate fungicides has been shown to provide a measure of control (Coffey, 1987; Giblin *et al.*, 2005; Kaiser *et al.*, 1997). There has also been research into the unconventional improvement of rootstocks to develop PRR resistant rootstock by methods such as somaclonal variation (Ben-Ya'acov & Michelson, 1995). To date this has not been successful.

© University of Pretoria

1.2.1 Avocado rootstock breeding

Avocado rootstock development is very important in sustainable crop production to overcome obstacles such as pathogens. As early as 1926 avocado researchers identified that the single most important problem was rootstock development and the success of the avocado industry was to be found therein. As time continued the clonal rootstock programmes gained footing and more people started to believe that clonal rootstocks would provide the answers to many problems such as growth, flowering habits, and most importantly disease resistance (Ben-Ya'acov & Michelson, 1995). The selection of disease tolerant rootstocks has become routine in the avocado industry since control methods of Phytophthora root rot are proving to be ineffective.

A very unique breeding programme is established at Riverside in California USA. This breeding programme contains the largest rootstock collection in the world and is housed at the University of California, Riverside (Coffey & Guillemet, 1987). Rootstock breeding took hold in the 1950s when George A. Zentmyer started making collections of avocado varieties and related species with the primary purpose of selecting material that would possibly show promise of PRR resistance (Zentmyer, 1957). Sixty six collections were made from Mexico, Guatemala, El Salvador, Honduras, Costa Rica, Colombia, Ecuador, Peru, Chile, Argentina, Brazil, Venezuela and Puerto Rico with some *Persea* species showing high resistance to PRR. These species could however not be grafted to *P. americana* fruiting cultivars and it was evident that fruit production is not only dependent on resistance but also on the scion/rootstock interaction. The variability of rootstock performance was the basis behind the breeding and selection process. This variability was dependant on climatic and environmental conditions too. This was demonstrated in the 1990s when the specificity of avocado cultivars to different environments was proved when the Sheppard rootstock showed the greatest sensitivity to temperature at

© University of Pretoria

flowering when tested across Australia (Whiley *et al.*, 1990). In 1956 the clonal rootstock Duke 7 was first identified as having moderate resistance to PRR and was subject to further investigation. In 1985 a study confirmed that two rootstocks, namely Duke 7 and G6 were partially resistant to *P. cinnamomi*. In 1990 tolerance to PRR was demonstrated in four other rootstocks in addition to Duke 7, these were Thomas, Martin Grande, Barr Duke and the D9 rootstocks (Gabor *et al.*, 1990; Zentmyer & Thorn, 1956). Duke 7 was the most tolerant rootstock on the market for close to 50 years and was deemed to be the commercial standard.

Production of clonal rootstocks in South Africa began long before the success of Dusa®. In the late 19th century the first avocado trees were brought to South Africa (Kwazulu Natal) and were probably of the West Indian race (Whiley *et al.*, 2002). The main production areas in South Africa are Mpumalanga and Kwazulu Natal (Ben-Ya'acov & Michelson, 1995) as well as the Limpopo province. Rootstock breeding programmes started with Westfalia Technological Services (WTS) embarking on rootstock development in the 1970s and the ARC-Institute for Tropical and Subtropical crops in 1992 to develop rootstocks resistant to PRR (Bijzet, 1999; Köhne, 2005). Due to the diligence of WTS, they have been able to produce their own rootstocks that are best suited for local growing conditions. This nullifies the need to bring in rootstocks that are developed in other countries and may not be that well suited for South African conditions (Ben-Ya'acov & Michelson, 1995).

Rootstock selection begins with open pollination between 20 trees in the breeding block at Westfalia. Seedlings from this block are exposed to PRR in a mist bed inoculated with *P. cinnamomi.* The surviving seedlings are then multiplied clonally and re-tested in the mistbed. These survivors are then grafted to the Hass scion and then subject to field testing (Kremer-Köhne & Mukhumo, 2003). *Persea indica* is used as an indicator plant for pathogenicity when testing in the mist bed due to its high sensitivity to *P. cinnamomi* (Ben-Ya'acov & Michelson,

© University of Pretoria

1995). Intensive trials and continuous testing bore fruit for WTS when Dusa® showed great potential after 20 years of development (Menge, 1999). Dusa® has been experimentally proven to be more tolerant and more productive than Duke 7 (Kremer-Köhne & Mukhumo, 2003; Menge, 1999) and has now displaced Duke 7 as the commercial standard in South Africa after its commercial release in 2001(Wolstenholme, 2003). Dusa® is currently being extensively used in South Africa and Australia and has shown promise in other countries such as Israel and in California in the United States of America.

There are various problems associated with rootstock breeding such as time and energy expenditure. Screening for resistance to *Phytophthora* and other traits is a lengthy process taking eight weeks to grow the avocado seedling, eight weeks to assess resistance to the pathogen in mistbed trials, followed by another eight weeks to retest. Field testing finally takes place over a period of 5-10 years (Gabor & Coffey, 1991). The success rate of rootstock selection is low with over 90 percent of selections failing due to chance selection or incompatibility between rootstock and scion resulting in reduced yields (Coffey & Guillemet, 1987). Rootstock screening and development is a time consuming and expensive process with little understanding of the *P. cinnamomi/*avocado dynamics or pathogen infection mechanisms.

Coupled to the economical problems are the problems of chance, incompatibility between scion and rootstock, developing tolerance to fungicides and the molecular hurdles involved in resistance rootstock breeding. There is always the aspect of a chance selection for a rootstock for tolerance to PRR with the selection performing badly in the field or when grafted to the specific fruiting scions. The high amount of heterozygosity of the angiosperm contributes to this problem by making it difficult to predict offspring performance (Davis *et al.*, 1998; Mhameed & Sharon, 1996). Recently, the need for enhanced tolerance breeding has been highlighted by a

© University of Pretoria

study showing a build up of phosphite resistance in *P. cinnamomi* that occurs with the prolonged usage of fungicides (Dobrowolski *et al.*, 2008).

1.2.2 Avocado Genomics

Phytophthora americana is a diploid angiosperm consisting of 24 chromosomes with approximately 8.83 x10^8 bp (bp) (Chanderbali *et al.*, 2008; Sharon *et al.*, 1997). Current data suggests that the avocado like other species in the family *Lauraceae* have undergone a genome duplication event which owes to the gene duplications seen in the crop (Chanderbali *et al.*, 2008). There is not much molecular information available for avocado and the avocado genome has not to date been released. Currently there are 16558 ESTs available on the NCBI for avocado however, these ESTs originate from fruiting and flowering organs of scions. There are 6183 unigenes available on the Plant Genome Network (PGN, http://pgn.cornell.edu/) but these sequence origins are from flower buds. Many of the accessions are duplications of house-keeping genes and tell us nothing about the genetic differences, variability in gene expression or PRR defence-related genes of the rootstocks.

The few genes that have been identified and studied thus far are from the fruiting scion of the avocado. Genes such as 9-cis-epoxycarotenoid dioxygenase (Chernys & Zeevaart, 2000), a cellullase gene member (Cass *et al.*, 1990) and the fruit ripening gene *Avoe3* (McGarvey *et al.*, 1992) have been identified with gene expression of the cellulase gene having been profiled in avocado fruit (Christoffersen *et al.*, 1984; Christoffersen *et al.*, 1982). The lack of molecular information of rootstocks or associated studies is not being addressed, while the ongoing research comprises of genetic relationship studies and the molecular characterization of the fruit. The greater part of molecular detail exists due to a continuous effort in marker

© University of Pretoria

development to assist in either elucidating genetic relationships (Acheampong *et al.*, 2008; Chanderbali *et al.*, 2008; Davis *et al.*, 1998; Fiedler *et al.*, 1998; Mhameed *et al.*, 1997), or crop improvement (Chen *et al.*, 2007; Clegg *et al.*, 1999; Lavi *et al.*, 1991) focussing on fruit quality.

1.2.3 Molecular markers in Avocado

One of the biotechnological inventions that greatly aid agriculture is the use of molecular markers in conventional breeding and selection of crops. A molecular marker allows selection based on a polymorphism either at the protein or DNA level. Examples of protein and DNA markers are isozymes and RFLPs respectively (Kumar, 1999). The application of markers has moved from random markers to more specific functional markers that are linked to the desired trait. This move has demonstrated the evolution from marker assisted breeding into genomics assisted breeding (Varshney *et al.*, 2005).

Marker development began in the 1980s and has been continuously evolving. From the inception of RFLP analysis there has been application of many different markers used in outlining some aspects of avocado biology such as phylogenetics, race identification and genetic linkage map creation. Some markers have advantages over others with regards to cost and amount of starting material (Table 2). Markers will be discussed in more detail in an order that represents their relative timeline of the application in the avocado industry.

© University of Pretoria

Table 2. Comparison between microsatellite, RAPD & RFLP molecular markers.

	Microsatellite	Randomly amplified polymorphic DNA	Restriction fragment length polymorphisms
Amount of DNA	Nanogram	Nanogram	Microgram
Inheritance of marker	Co dominant	Dominant/Recessive	Co dominant
Polymorphism	Highly	Moderately	Moderately
Development time/cost	High	Low/moderate	High
Cost per assay	Relatively low	Relatively low	Relatively high

Adapted from Clegg *et al.* (1999)

1.2.3.1 Isozymes

Isozymes are protein markers and have the ability to be used for differentiation between different individuals. This ability is due to the principal that allelic variation exists between many different proteins (Kumar, 1999).The first instance of molecular genetics being applied to avocado was in 1978. This application of isozymes was used to study single gene characters using the mesocarp of avocados. The development of isozymes was proposed to be used to study evolution and relationships of the different avocado races (Torres *et al.*, 1978).

© University of Pretoria

In 1980 isozymes from fruit and leaf origins were studied from over 100 cultivars based on the previous work done in 1978. Isozyme variation of peroxidase, malate dehydrogenase, leucine aminopeptidase and phosphoglucose mutase was studied revealing that isozyme data could be used to examine systematic and evolutionary aspects with new alleles being reported for some genes (Torres & Bergh, 1980). This study provided 12 genes as genetic markers which represented 6 enzyme systems and were specified by 37 co-dominant alleles.

Several drawbacks should be noted with regards to isozyme markers. There are a limited number of loci that can function as markers and for many species only 15-20 loci can be mapped. Only 40-50 reagent systems have been developed that permit the staining of a particular protein in a starch and not all reagent systems work efficiently with all plant species. Finally, there is a link between tissue variability in protein variation. The isozymes found in avocado had limited use because the polymorphic gene loci had limited genetic resolution (Clegg & Davis, 1989; Kobayashi *et al.*, 2000). Although they lacked the informative nature of today's marker systems, isozymes signified the beginning of marker development for use in avocado breeding (Clegg *et al.*, 1999). Before 1989 isozymes were the only source of genetic markers in avocado.

1.2.3.2 RFLPs

In 1989 the first use of RFLPs was reported. RFLP analysis is performed by fragmenting the DNA sample by restriction enzymes and the resulting fragments are separated according to their lengths by gel electrophoresis. An RFLP is a sequence of DNA that has a restriction site on each end with a target sequence in between. The target sequence binds to a probe by forming

© University of Pretoria

complementary bp (Kumar, 1999). The advantage of RFLPs is the large quantity of markers while the drawback is the low level of polymorphism (Lavi *et al.*, 1998).

Twenty one RFLP markers were used to screen 14 avocado cultivars. Seven multiple-copy sequences could possibly be used in cultivar identification. This was the first report of RFLP markers being used in avocado and that cultivars could be distinguished based on their multi-locus genotype (Clegg *et al.*, 1992).

RFLP analyses were conducted in 1998 for genealogical identification of cultivars (Davis *et al.*, 1998). Thirteen probes determined 14 genetic loci which were used to screen 36 cultivars. There were two common cultivars in this study and that of Mhameed *et al.* (1997) who used minisatellite markers. The conclusions found regarding their relationships were similar in both studies and grouped the cultivars into the three races- Guatemalan, West Indian and Mexican.

RFLP markers were also used to distinguish among breeding lines and again unravel genealogical relationships (Clegg *et al.*, 1999). Seed was collected by Coffey (1987) and three lines were labelled as G755A, B and C. G755A was identified as a cross between a known tolerant relative of avocado and *P. americana*. The hybrid showed root rot tolerance and the three seed lines were investigated using RFLP probes. The RFLP analysis revealed a unique genotype for each line. A routine assay was developed to distinguish between the three rootstocks to identify progeny with different levels of tolerance. When the relationships between 15 sources of avocado DNA were tested, it was conclusive that the RFLP pattern was distinct for each. The cultivars grouped into the three botanical races and demonstrated the power of the method to identify unknown cultivars (Coffey & Guillemet, 1987).

© University of Pretoria

1.2.3.3 RAPD

Random Amplified Polymorphic DNA (RAPD) markers are DNA fragments generated from random amplification of genomic DNA. These markers use short random primers for amplification and the polymorphism is generated by the primers ability to anneal to the sequence (Kumar, 1999). The primers ability to anneal is affected by mutations in the genome.

The advantage of using RAPD markers is that no knowledge of target DNA sequence is required. The number of amplification products (banding pattern) is directly related to the number and orientation of the sequences that are complementary to the primer in the genome (Kobayashi *et al.*, 2000). This molecular marker has caught the attention of those involved in the field of applied plant breeding. The drawbacks of RAPDs are; the low level of polymorphism alongside the irreproducibility of the banding pattern and the dominant mode of inheritance. Its advantage however, is the low cost, ease of work and the fact that it is PCR based (Durban *et al.*, 2011).

RAPD markers were assessed for their use identifying different races in avocado (Chang *et al.*, 1993). Sixteen samples from avocado germplasm collections in Israel were tested that represented the 3 races. Of the 71 primers tested, 22 informative primers were identified and used to generate 111 polymorphic DNA fragments. Of the 22 primers the authors identified three primers which could be used to distinguish between all the avocado accessions. The RAPD markers proved to be successful in identifying between different avocado cultivars and suitable for studying genetic relationships due to their high polymorphic nature.

Outcrossing in avocado orchards was studied in 1999 using RAPD markers in orchards of Southern California (Clegg *et al.*, 1999). Over 300 primers were tested on 4 cultivars (Hass, Fuerte, Bacon and Zutano) to identify which pollen source was most efficiently crossed with the

© University of Pretoria

Hass cultivar. Seven different primers revealed differential banding across the cultivars and identified that Fuerte pollen was the most effective pollen source for maternal trees.

Outcrossing in avocado orchards was again studied in 2000 using RAPD markers on the same cultivars as Clegg *et al.* (1999) (Kobayashi *et al.*, 2000). Outcrossing rates, cultivar identification and estimation of the relationship between fruit yield and outcrossing were some of the objectives hoped to be achieved. The results from this study were similar to those achieved prior using RFLP markers and micro/minisatellite markers. There was a weak correlation between outcrossing rate and yield. It was shown that outcrossing was dependant on the distance from a pollen source and that self-fertilization was responsible for a fairly large amount of the fruit set in California (Kobayashi *et al.*, 2000).

1.2.3.4 VNTRs

VNTRs (Variable Number of Tandem Repeat) are locations in a genome where a short nucleotide sequence is organized as a tandem repeat. These tandem repeat regions are called minisatellites or microsatellites depending on the length of the repeat region (Durban *et al.*, 2011). Polymorphisms in these regions due to allelic variation in the number of tandem repeats, and these repeats often show variations in length between individuals.

There are two classes of VNTRs, DFPs (DNA fingerprints) and SSRs (simple sequence repeats). DFPs are similar to RFLPs because restriction enzymes are employed to cut the sequences flanking the repeat and VNTR markers are usually genotyped using Southern blotting. SSRs are abundant in plants and are flanked by conserved regions that enable primers to be developed to amplify the region. The main advantage of VNTRs is that they are highly

© University of Pretoria

polymorphic and PCR based (Durban *et al.*, 2011). Its only disadvantage is the cost but this has been reduced considerably since its inception.

In a study by Lavi *et al.* (1991) four DFPs were used to investigate 14 cultivars and each cultivar could be identified using its DFP pattern. High levels of polymorphic banding patterns were achieved using three minisatellite probes. DFP analysis done on avocado rootstocks using mini- and microsatellite probes resulted in each cultivar being identified by its DFP pattern and each race also being characterised by a specific DFP pattern. The study suggested that the DNA fingerprints would be useful for breeding for economically important traits, allowing the rapid screening of seedlings in avocado breeding programmes (Lavi *et al.*, 1991).

In 1994 another VNTR marker was identified as an alternative to RFLP markers, namely SSRs. SSRs are microsatellites of repeating sequences of 1-6 bp of DNA. The basis of SSRs is the same as that of VNTRs. These microsatellites are typically neutral and co-dominant. They are used as molecular markers in genetics, kinship, and population and can also be used to study gene duplication or deletion. These SSRs were shown to be highly polymorphic and very abundant in avocado (Lavi *et al.*, 1994). The study used four dinucleotide probes to probe repeat lengths in 12 cultivars. The Hass cultivar was used as the genomic library to select SSR probes. The study indicated that there was one SSR every 13.1 kbp (kilo basepair) and that cultivar identification was successful.

The heterozygosity level in avocado was measured in 1996 using DFP and SSR markers (Mhameed & Sharon, 1996). DNA fingerprinting was done on 15 progeny using a single probe and 50 progeny were analysed at 55 SSR loci. Both DFP and SSR markers revealed a high level of heterozygosity in the avocado genome although SSR heterozygosity was lower than that of DFP heterozygosity. The high level of heterozygosity is an advantage to the plant due to genetic diversity but is a disadvantage for the breeder due to the difficulty in predicting offspring

© University of Pretoria

performance (Davis *et al.*, 1998; Mhameed & Sharon, 1996). DFP markers are well suited for identification purposes and SSR markers are more suitable for linkage analysis based on heterozygosity. The authors concluded that ultimately the choice of marker will depend on the objective of the research project.

In 1997 DFP markers were used to assess the genetic relationship between 32 avocado cultivars (Mhameed *et al.*, 1997). Thirty-two cultivars were assessed with 24 being allotted to one of three races. The study reached the same conclusion as that of Lavi *et al.* (1991) in separating West-Indian, Guatemalan and Mexican races.

The genealogical relationships of avocado were once again studied in 2003 (Ashworth & Clegg, 2003). This time the authors noted that the previous studies were lacking statistical data to support their findings and measures of statistical support were applied to the data in the form of bootstrap values in an un-rooted neighbour joining tree. Ten dinucleotide and 15 trinucleotide markers (25 markers in total) were used to analyse 37 avocado samples. No conclusive result was achieved in this study due to the fact that bootstrap values were low across the entire study. The low bootstrap values could partially be explained by the possibility of hybrid samples being analysed. The samples analysed did however associate into three groups that represent the three botanical races, albeit being loosely grouped based on bootstrap values.

In 2003, another study involved the development of microsatellite markers (127) to aid in reducing the time needed for avocado breeding (Ashworth *et al.*, 2004). Thirty-seven avocado genotypes were typed at 25 loci. This revealed 204 alleles comprising of di- and trinucleotide loci. The study indicated that dinucleotide repeats yield more loci, development of trinucleotide repeats are more costly due to sequencing costs and that dinucleotide repeats are more abundant. The authors also suggested that the use of a single cultivar for the development of a

© University of Pretoria

microsatellite library is too narrow for broad spectrum marker development due to the genetic diversity present in avocado.

The genetic diversity of a germplasm collection in the USA was also tested using microsatellites and statistical tools to support findings (Schnell *et al.*, 2003). Fourteen microsatellite markers were used to analyse a larger sample set of 224 accessions. The authors found an average of 18.8 alleles per locus but no race specific markers were identified although the majority of the accessions could be grouped under the three botanical races, yet there were some anomalies present. Their findings indicated that the germplasm resource in Miami was dominated by the Guatemalan and Mexican races with the least amount of diversity belonging to the West Indian race. Since tolerance to PRR had previously been identified in the West Indian race it has been suggested that more West Indian genetic diversity needs to be introduced into the germplasm resource.

SSR markers were used once again to test genetic diversity in a germplasm collection, this time being applied to Spanish genotypes (Alcaraz & Hormaza, 2007). A total of 75 samples (29 Spanish and 46 from other countries) were analyzed using 16 markers previously developed by Sharon *et al.* (1997) and Ashworth *et al.* (2004). The study identified 156 fragments with similar sizes as indicated in other studies. Overall the data agreed on many levels with that of previous studies in grouping of cultivars, marker size differences and levels of inbreeding. The study could not identify the racial origin of the different cultivars, this is suspected to be caused by the fact that all the samples are in fact hybrids and a deeper study may need to be conducted comprising of more markers.

SSR markers were used to study the genetic diversity of avocados in Ghana (Acheampong *et al.*, 2008). This study used 12 SSR markers to study 172 samples and generated 53 alleles. It concluded that the majority of Ghanaian avocados belong to the West Indian race and that more

© University of Pretoria

genetic diversity needs to be introduced to improve the industry. The study also had some difficulty in distinguishing between certain samples; this could be due to the lack in marker diversity as indicated by Ashworth *et al.* (2004). The study also showed clustering of the races into their three different groups.

1.2.3.5 Avocado genetic linkage map

An important application of DNA markers is linkage analysis whereby traits are mapped to markers. A genetic linkage map is not a physical map and does not use measures of physical distance. It is a genetic map that shows the position of its known genes or genetic markers relative to each other in terms of recombination frequency (Collard *et al*., 2005). The greater the frequency of recombination between two genetic markers, the farther apart they are assumed to be. DNA sequences such as microsatellites or those generating RFLPs have been used to create genetic linkage maps.

The first genetic linkage map was created for avocado comprising of different types of markers (Sharon *et al*., 1997). The map was aimed at aiding in QTL (quantitative trait loci) identification for the breeding of avocados with desired traits. Fifty polymorphic SSR loci, 17 RAPD and 23 mini-satellite DFP bands were used to construct the avocado genetic map. From the 90 SSR, RAPD and DFP markers 56 markers remain unlinked. Twelve linkage groups were detected covering approximately 40% of the genome. It was concluded that for avocado and other sub-tropical fruit trees SSR markers would be the most desirable marker system for the purpose of linkage map development since they were polymorphic, abundant and locus-specific and overcame the shortcomings of RAPD markers that were not that informative and DFP markers that are not locus specific.

© University of Pretoria

1.2.3.6 QTLs

Many agriculturally important traits are controlled by many genes and are known as quantitative traits. The regions within genomes that contain genes associated with a particular quantitative trait are known as quantitative trait loci (QTLs) (Collard *et al.*, 2005). QTL analysis links a marker to the trait of interest. If the trait and marker are closely linked then the presence of the marker is usually interpreted as the presence of the trait (Kumar, 1999). The main application for a linkage map is to aid in detecting the loci by association of QTLs with markers.

The genetic linkage map created by Sharon *et al.* (1997) was used to identify these associations in a study in 1998. The study was conducted to screen associations between markers and eight avocado fruit traits (Sharon *et al.*, 1998). Samples were taken from the plants, leaves and traits that were studied were fruit weight, skin colour, skin gloss, skin thickness, skin peeling, seed size, flesh fibres and taste. RAPD markers and SSR markers were used to detect QTLs for the traits, a total of 90 markers were chosen. The study revealed that there were many problems associated with the identification of QTLs, specifically that the statistical significance threshold criteria ((P value) of ≤ 0.01) was not satisfied for some traits. From the traits that conformed to the statistical threshold there was no association between fruit weight and any marker loci or between skin colour and any marker but two QTLs were identified. The QTL controlling fibres in the flesh of the fruit was linked to the marker AVA04 (SSR marker) with the QTL controlling skin gloss linked to marker RAPD16 (RAPD marker).

The growth rate of avocado was studied by trying to associate a few QTLs to microsatellite markers (Chen *et al.*, 2007). Two hundred and four genotypes were selected to be used in this study using 10 microsatellite markers that were developed by Ashworth *et al.* (2004). The traits that were studied over a four year period were tree height, average canopy diameter, average

© University of Pretoria

tree trunk diameter, flower abundance and fruit set. Most of the traits had high correlations between each other except that fruit set had a medium correlation and flower abundance showed no correlation to the growth rate. No genotype-by-location effect was noted for the growth rate with the authors concluding that different pollen parents have an impact on growth rates because canopy differences were also noted when 'Gwen' cultivars were crossed with other cultivars. The information generated in the study on avocado growth rates indicated that selection for high growth rate trees may lead to enhanced fruit set. This study indicated the fundamental application of the markers developed prior to its execution.

It is clear from afore mentioned studies that there is an ongoing effort to initialize the use of molecular tools in avocado breeding and to develop new tools that will reduce the time and effort involved in the process. The development of RFLPs saw the beginning of DNA marker identification. To date the only QTL study conducted was on growth rate, flower abundance and fruit set with markers being identified for various traits such as fruit weight, seed size, skin properties and harvest duration. All of these studies have been conducted on fruiting and flowering characteristics (Chanderbali *et al.*, 2008; Chen *et al.*, 2007).

Markers could assist in the fight against PRR. In the study by Clegg *et al.* (1999) it was shown that RFLP markers were able to identify different rootstock sources and possibly be used to select rootstocks with different tolerance levels to PRR. The newest development in the fight against PRR was the development of 70 microsatellite markers developed from over 8000 ESTs in the hope of aiding in marker assisted breeding against PRR (Borrone *et al.*, 2007). Seventy loci were identified from 24 avocado samples with an average of 7.1 alleles per locus. The primers were tested across various crosses of avocado but the study was only aimed at increasing the number of markers available. Thus no further result was generated accept the

© University of Pretoria

applicability of the markers but the downfall of this study is that the material used for the marker development was flower buds.

Since screening for resistance to *Phytophthora* and other traits is a lengthy process taking up to 10 years there is a need for development of faster and more efficient methods of crop development. Resistance in rootstocks was discovered decades ago (Kellam & Coffey, 1985) but no biological basis for resistance has been elucidated. Even with current developments and technologies there is no molecular data of resistance to *Phytophthora cinnamomi* exhibited in rootstocks. Most if not all molecular studies discussed above are concerned with the identification of cultivars and not with identifying disease resistance markers.

Research on defence responses in avocado is scant with some defence responses being studied in the avocado root and *Arabidopsis thaliana* recently. The study conducted on avocado roots investigated antioxidant reactions and specifically the role of nitric oxide in the response to *P. cinnamomi.* The study did however, not include any gene discovery or transcriptional analysis (García-Pineda *et al.*, 2010). Defence responses related to *P. cinnamomi* were studied in *Arabidopsis* (Rookes *et al.*, 2008). The responses observed were similar to other *Arabidopsis* defence responses such as reactive oxygen species (ROS) induction, hyper-sensitive response (HR) activation, and lignin synthesis and callose production. Macroscopic changes such as callose production have been observed during *P. cinnamomi* in previous studies (Hinch & Clarke, 1982) but the study by Rookes *et al.* (2008) gave the first gene level information on *P. cinnamomi* interactions. They observed that the general host defence response involved the activation of ethylene and jasmonic acid pathways rather than the salicylic acid pathway. A somewhat different result was obtained by García-Pineda *et al.* (2010). The authors indicated that salicylic acid is a major inhibitor of pathogen colonisation. These discrepancies illustrate the

© University of Pretoria

need for molecular studies to be conducted on the host plant since *Arabidopsis* is a non-host of the pathogen.

The feasibility of using molecular markers for avocado improvement was established in 1989 (Clegg & Davis, 1989). Avocado breeding programmes will benefit from marker assisted breeding due to the time consuming and expensive screening of seedlings. Recently the use of markers in improvement of avocado breeding had been exemplified (Clegg *et al.*, 1999) although the successful application of molecular markers in avocado has only gone as far as to differentiate between the different races and cultivars using RFLPs, RAPDs and VNTRs. There is a need for the development of more tightly linked, specific and informative markers specifically from avocado rootstocks for use in the fight against PRR.

Sequencing of agriculturally important crops such as avocado, will allow identification of molecular markers and gene identification. Marker development has come a long way since its inception in the early 1980s and current marker systems are based on an indirect method of identifying traits and do not use the specific sequence of interest (gene) (Durban *et al.*, 2011). The new technology of high-throughput sequencing platforms shows promise in improving marker development and molecular characterization of this economically important crop.

1.3 High-throughput sequencing platforms

High-throughput sequencing has improved greatly on the methods of the past in generating a large amount of different sequence reads because of increased throughput and base calling accuracy in a single run (Scheibye-Alsing *et al.*, 2009), this reduces the cost of sequencing and also increases the depth due to the number of redundant reads generated. Prior to high-throughput sequencing the basic procedure that was used in transcript sequencing was to

© University of Pretoria

create cDNA and clone it into appropriate vectors, creating a library of clones. These were then individually sequenced (using Sanger sequencing) and the sequences were aligned, which is a laborious and time consuming method. For the sequencing of genomes methods such as shotgun sequencing, hierarchical sequencing, mixed strategy sequencing and ESTs were employed. Previous methods were not well suited to analyse large amounts of sequence data and were coupled to high costs and excessive time which served as prohibitions when large scale genome or transcriptome studies were considered (Scheibye-Alsing *et al.*, 2009).

Next generation sequencing (NGS) has surpassed previous sequencing techniques for genome and transcriptome analysis and allowed many researches to tackle questions on a larger scale. Large scale genome and transcriptome analysis became possible with the advent of microarray technology and recently sequencing platforms have granted the ability to produce large amounts of data in far less time than before (Metzker, 2010). The advent of NGS platforms such as Illumina, GS FLX-454 and SOLiD have lead to a more efficient procedure of analysis while also increasing the depth of coverage (Collins *et al.*, 2008; Metzker, 2010). What makes NGS platforms ideal for mRNA analysis is the elimination of bacterial cloning (Simon *et al.*, 2009) and the lower cost involved in sequencing when opposed to that of capillary sequencing (Wall *et al.*, 2009). Biases of instability in the cloning process of previous methods are avoided with pyrosequencing of cDNA transcripts (Weber *et al.*, 2007). The generic terminology used to describe these NGS platforms is high-throughput sequencing (HTS) since what was next generation is actually the now generation sequencing platforms.

HTS technologies rely on two principals: immobilisation of the nucleotide sequence for interrogation onto a substrate and adding a base that results in a visible change that can be quantified and recorded as an image at each cycle, hence the terminology ‘sequencing by synthesis’ of each fragment (Scheibye-Alsing *et al.*, 2009). There are two methods for each

© University of Pretoria

principal of preparing samples for sequencing by these platforms. With regards to immobilisation, the nucleotides are fixed to a glass slide or to streptavidin beads, with regards to amplification templates are either clonally amplified (emulsion PCR for the 454 GS FLX platform) or they are single molecule templates (Helioscope by Helicos) (Metzker, 2010). The specifics of each principal differ between technologies and ultimately, the sequence data produced is affected by the specific combination of template preparation, sequencing, imaging and data analysis.

1.3.1 SOLiD

The SOLiD system by Applied Biosystems (http://www.appliedbiosystems.com) employs a complex procedure with fluorescently labelled probes reading different lengths of the fragment by selective cleaving in cycles. The system is similar to a combination of the 454 and Illumina platforms. The ligation of eight-base degenerate probes to the fragments in an n-1 fashion through each cycle of ligation allows the interrogation of each base twice and gives the platform its uniqueness (Holt & Jones, 2008). Emulsion PCR is followed by binding of beads to a glass slide and then placed into a fluidics cassette. The adapter specific primers and fluorescently labelled nucleotides are then added and sequence data is generated using ligase-mediated sequencing. Four-colour oligonucleotides compete for ligation with each oligonucleotide that are made of three universal bases and two fixed bases. This means that the bases being queried are offset by three nucleotides from the point of ligation. Another difference in this sequencing technology is the use of ‘colour space’ to interpret the generated data (Metzker, 2010).

The data is called in colour rather than bp with four different colours being used for the four nucleotides; however, the downfall is that all further analysis should be performed in colour

© University of Pretoria

space to avoid complications and errors. Sequencing errors may occur when colour-space sequence is translated into text. The error will result in a frame-shift and all data from the erroneous base will be incorrect (Holt & Jones, 2008). Another disadvantage is the lengthy time taken to produce data which may take 3-5 days (Rothberg & Leamon, 2008). The advantage of this is the increased accuracy of the method and the generation of 50 bp reads with a total of 30 Gb data being generated (Metzker, 2010).

1.3.2 Illumina

Solexa sequencing now owned by Illumina (http://www.illumina.com) involves immobilizing fragments of adapter-tagged DNA and sequencing using removable fluorescently labelled terminators. Sample preparation is initiated with the linking of adapters to cDNA and loading onto a flowcell using a microfluidic cluster station which creates around 100 – 200 million clusters (Metzker, 2010). The flowcell is coated with covalently linked oligonucleotides. The solid phase amplification takes place with the free adapter end bending to bind to an oligonucleotide creating a bridge, amplification then creates double stranded bridges. The process is repeated numerous times to form clusters of the same fragment, thus achieving clonal clusters of around 1000 copies. After the solid phase amplification the flowcell is moved to a fluidics cassette in which the templates are sequenced using four differently coloured nucleotides (fluorescently labelled). After the incorporation of a coloured reversible terminator, the sequence is read and the terminator released by unblocking the 3' end to allow for the incorporation of the next base (Metzker, 2010; Morozova & Marra, 2008). The innovation of the system is the template amplification and four-color reversible, terminators (Holt & Jones, 2008). The platform generates 75 – 100 bp fragments generating >18 gigabases (Gb) of sequence

© University of Pretoria

information over four days. More data (up to 35 Gb) can be generated using the newly developed paired-end component whereby sequencing occurs in both the 3' and 5' directions from a single template (Metzker, 2010; Scheibye-Alsing *et al.*, 2009; Simon *et al.*, 2009).

The RNA-seq approach is used to discover new exons or genes by sequencing cDNA. Poly (A) RNA is selectively fragmented to ~200bp, converted to cDNA by random priming and adapters are added. It is preferable to fragment then random prime to synthesize cDNA and then add adapters at the single stranded stage. The biggest limitation is the short sequence reads (Simon *et al.*, 2009). The RNA-seq approach is sufficient for expression profiling but not for empirical gene annotation. This is because of the modest sequence coverage over the entire gene length and may be limited by the absence of a reference genome (Priest *et al.*, 2010).

1.3.3 454 pyrosequencing

The 454 pyrosequencing technology has the advantage of longer sequence length with regard to the identification of novel transcripts (Simon *et al.*, 2009). The platform emerged with the GS20 which generated read lengths of around 100 bp but then was soon upgraded to the GS-FLX platform that could generate read lengths of ~400 bp resulting in information of about 20 million bp (Margulies *et al.*, 2005; Morozova & Marra, 2008; Ronaghi *et al.*, 1998; Simon *et al.*, 2009). In 2008 Roche launched the Titanium series of reagents which could generate 400 – 600 million bp per instrument run with read lengths of 400 - 600 bp, the new GS FLX Titanium XL+ produces reads of up to 1000 bp but on average around 700 bp with 85% of the reads being above 500 bp. The current platforms' throughput is around 700 Mb. This indicates that although the amount of data capable of being generated has gone up substantially, the cost has not increased dramatically. Where it was previously required to cover most of the transcriptome with

© University of Pretoria

2 GS20 runs it is now possible to use a single run of the GS-FLX with titanium reagents since there has been an approximate 25 fold increase in the data generated. When compared to traditional EST sequencing, pyrosequencing discovers three times more genes and covers twice as much sequence (Weber *et al.*, 2007).

The GS FLX-454 system from Roche (http://www.454.com) employs synthesis and sequencing reactions within droplets of emulsion, which act as individual amplification reactors utilizing pyrosequencing. DNA is sheared by sonication or nebulisation and isolated in these emulsion droplets to which adapters are bound and captured on individual streptavidin beads. The emulsion amplification step generates template for the sequencing and detection phase to be carried out (Metzker, 2010). Single bases are added one base at a time and the inserted base is interrogated. Before the next nucleotide is added a wash step removes unbound nucleotides. Detection is made possible by the pyrophosphates released during polymerase action, fuelling the luciferase reaction which occurs in open wells of fibre optic slides (PicoTitre Plate). The wells are big enough to allow a single bead to occupy it. The platform uses luciferase activity to visualize addition of bases. A camera connected to the wells records the sequencing process when photons are released by the reaction as nucleotides are washed over the wells.

This sequencing platform is very efficient in more than one regard. It reduces the amount of time required for library preparation, is more cost effective than prior sequencing methods, produces a large amount of data and can be used for multiple applications. Applications such as mutation detection, SNP genotyping, microbial genotyping, tag sequencing (ESTs) and gene identification (Ahmadian *et al.*, 2006). The platform also does not produce any substitution errors due to the sequential flow of nucleotides (Meyer *et al.*, 2009).

© University of Pretoria

The downfalls of this platform are the occurrence of insertions and deletions, failed sequences in wells, homopolymer sequencing problems and multiple copies of similar sequence on different beads. In the study by Weber *et al.* (2007) it was indicated that results obtained were unbiased but in another study however, it has been shown that the platform does show a 3' bias in non-normalized libraries (Wall *et al.*, 2009). Although the 454 platform does produce longer reads, it is around 30x more expensive than the Illumina platform (Holt & Jones, 2008; Mardis, 2007; Simon *et al.*, 2009).

When the Illumina platform was compared to GS20, GS-FLX and traditional EST sequencing it was shown that Illumina generates better transcriptome coverage when taking cost into consideration with the main drawback being the assembly of the short reads into full length sequences (Wall *et al.*, 2009). The platform also outperformed the other methods when the number of genes tagged was assessed. The Illumina platform experiences difficulty when it comes to gene coverage and singleton reads due to its short reads. Some of the difficulties encountered are that more sequencing is required for the same genome coverage as 454 (Rothberg & Leamon, 2008) and there is an inherent difficulty in repeat assembly due to the short read lengths of under 100 bp.

Each of the technologies described above has a negative aspect whether it is cost, speed or accuracy (Coombs, 2008). One of the problems with these approaches is the large amount of starting material needed (Simon *et al.*, 2009). The short reads also pose problems for assembly of sequence repeats (Morozova & Marra, 2008; Trick *et al.*, 2009) especially with the Illumina and SOLiD platforms. The best results are obtained using more than one sequencing technology as they may solve some of the individual problems encountered (Scheibye-Alsing *et al.*, 2009 ; Simon *et al.*, 2009; Wall *et al.*, 2009).

© University of Pretoria

Table 3. Overview of high-throughput sequencing platforms.

Company	Platform	$ (instrument)	$ (reagent costs)	Data output per run (Gb)	Read length (nucleotides)
Applied Biosystems	Bead-based clonal ligation DNA sequencing	$595,000	$3,400	30 Gb	50 nt
Helicos	MPS single molecule sequencing	$999,000	$18,000	37 Gb	25-50 nt
Illumina	Sequencing by synthesis	$600,000	$4 –6,000	18 –35 Gb	36 -100 nt
Pacific Biosciences	Single molecule real time sequencing by synthesis	$600,000 (anticipated)	$99	72 Mb	580 – 2,800 nt
Roche (454)	Massively parallel pyrosequencing by synthesis	$500,000	$5,000	0.45 Gb	400 – 600 nt

(Adapted from Coombs, 2008 , Holt & Jones, 2008)

© University of Pretoria

1.3.4 Applications

New sequencing technologies have multiple applications which are afforded by their high-throughput nature. Tag-based studies were one of the earliest applications of high-throughput sequencing technologies but has now even replaced EST sequencing due to their inherent depth (Simon *et al.*, 2009). Due to their 'deep sequencing' nature these platforms can be used for a variety of applications such as genome-wide chromatin immunoprecipitation sequencing (ChIP-seq), SNP analyses, small non-coding RNA discovery, metagenomics and transcript profiling (Coombs, 2008; Metzker, 2010; Morozova & Marra, 2008). Other studies have also been performed such as mapping of transcription factor binding sites (Johnson *et al.*, 2007). Complementary DNA microarray slides could also be developed from the sequencing data generated by these high-throughput platforms.

Studies ranging from the use of catfish ESTs to the use of 454 and Sanger sequence data from *Melitaea cinxia* have been used to create microarray slides (Li *et al.*, 2007; Vera *et al.*, 2008). Microarray probes were also developed by Meyer *et al.* (2009) from the sequence data generated for *Acropora millepora* coral. High-throughput sequencing platforms have also been particularly effective in generating SNP (single nucleotide polymorphisms) markers. This can be applied to studies of the *Phytophthora*-avocado interaction and speed up gene identification.

The 'seq-based' methods offered by high-throughput platforms have impinged on the microarray market (Metzker, 2010). High-throughput sequencing technology contends with microarrays for RNA transcript analysis due to the latter having confinements such as background noise and cross hybridization of a single RNA species to multiple probes. With the development of 'deep sequencing' methods around 2005 (Wall *et al.*, 2009) it is now possible to more accurately identify and characterise genes of interest by studying the transcriptome of plants.

© University of Pretoria

1.3.4.1 Transcriptome analysis

The transcriptome is the complete set of RNA transcripts produced by the genome at any one time and is very dynamic. It shows much variation with regards to cell type and environmental conditions i.e. having different patterns of gene expression. Transcriptomics is the study of the transcriptome and a global way of looking at gene expression patterns. It has allowed us to realise the full scope of RNA transcript changes that take place during biotic and abiotic influences on an organism (Fernandes *et al.*, 2008). This allows researchers to closely monitor gene regulation and uncover gene function, sequencing the transcriptome has also enabled functional studies on global gene expression (Vera *et al.*, 2008). Transcriptomics was developed in order to facilitate genome research and adds to a number of other resources such as linkage maps, physical maps and ESTs (Li *et al.*, 2007).

HTS data consists of more than just nucleotide sequence reads. The data generated by these high-throughput sequencing platforms comes in the form of DNA sequence reads of varying lengths and in great magnitude (Table 3). Along with the sequence data, files that contain the sequence run information, the quality scores of the data and other technical information are fed into the computational analysis. Analysis of this data requires great computing power and skilled knowledge of computer scripting (Jackson *et al.*, 2009). The generated sequence reads are assembled via identifying overlapping reads and joining these reads that have significant overlapping regions resulting in contig generation. The tools for mapping of reads generated by these platforms such as MAQ, SOAP, SHRiMP, VELVET, Exonerate and ELAND are used to align the sequence reads to reference genomes (Metzker, 2010). The BLAST tool can be used for overlap detection with various computational algorithms aiding in the assembly but cannot be solely utilized for alignment because it would take more than ten years to align 40 million, 36 base pair reads (Holt & Jones, 2008; Scheibye-Alsing *et al.*, 2009). Various high throughput

© University of Pretoria

platforms are available but the dominant platforms for sequencing that are currently in use are 454, Illumina and SOLiD.

A study by Weber *et al.* (2007) indicated that although thousands of ESTs are present in databases, pyrosequencing of the *Arabidopsis* transcriptome generated sequences that were not represented in any database. This gives an insight as to the depth achievable by this high-throughput platform. The study also indicated that full coverage of transcripts could not be significantly increased with random priming for cDNA library generation. Results obtained were unbiased, no sequence was lost during cDNA library generation due to size exclusion and the turnaround time for data generation was very rapid.

Wall *et al.* (2009) used the 454 GS20 platform to sequence an *Arabidopsis* cDNA library and generated a reasonably large number of reads (134, 781 in total) but with low coverage of the transcriptome. This was accomplished by using a half plate of GS20 (a full plate being capable of generating ~20 Mb of data). This study indicated the effectiveness of using the pyrosequencing platform for generating nearly complete genes in large numbers. It was however demonstrated in *de novo* sequencing of the coral transcriptome that the platform was indeed capable of generating ~100% transcriptome coverage (Meyer *et al.*, 2009).

For the *de novo* transcriptome sequencing of non-model organisms, the 454 pyrosequencing platform is the best option due to its long read lengths. Currently the platforms that could be used for transcriptome profiling generate short reads, such as Illumina and Helicos, and are cheaper than those using longer reads (454) but at the same time they pose challenges for *de novo* assembly applications (Sorber *et al.*, 2008). These short reads make it difficult to unequivocally align sequence data to a reference genome but due to the large amount of redundant sequence generated it is a tolerable trade-off (Simon *et al.*, 2009). If no other closely related organism's genome sequence is available for alignment then the use of 454

© University of Pretoria

pyrosequencing is the more favourable option. For *de novo* sequencing of transcriptomes it is better to use longer reads but the use of paired-end reads will double the amount of data generated and improve on sequence assembly (Farrer *et al.*, 2009; Morozova & Marra, 2008).

1.3.4.2 *De novo* sequencing

The initial use of short-read sequencing was for matching data to an analogous reference genome to identify differences such as, methylation, SNP identification and finally gene expression. *De novo* methods of transcriptome sequencing have been attempted before (Collins *et al.*, 2008; Vera *et al.*, 2008) (Figure 1). This is useful for organisms whose genomes have not yet been sequenced. Many of the economically important crops have not been sequenced and these methods will prove useful for their characterization and improvement. The efficiency of these techniques for gene discovery stems from two innovations: Firstly, by doing away with the labour intensive cloning of cDNA into vectors, and secondly, every cDNA sequence is protein coding along with splice variants being detectable and non-coding regions are avoided, (Weber *et al.*, 2007).

Collins *et al.,* (2008) used the Illumina sequencing platform for the transcriptome analysis of a non-model organism (*Pachycladon enyssii)* according to the outline described in figure 1. To manage data, the MySQL database was used and mapping was done using the Illumina Genome Analyser Pipeline software; ELAND. After optimisation of mapping, *de novo* assembly of contigs was carried out using version 0.5 of VELVET software. This research evidenced the use of a distantly related reference genome for mapping and *de novo* construction of a non-model organism's transcriptome.

© University of Pretoria

A study was conducted by Farrer *et al.* (2009) to investigate *de novo* assembly of the *Pseudomonas syringae* B728a genome using the paired-end module offered by Illumina with VELVET and EDNA software being used for assembly. The study indicated that 3% of the reference genome was not sequenced and related to non-coding RNAs or mobile genetic elements, but >90% of protein-coding genes were assembled with 100% accuracy. In conclusion the Illumina platform using the paired-end module is useful for *de novo* sequencing and also resulted in larger contig generation when using VELVET.

Trick *et al.* (2009) performed SNP analysis in *Brassica napus* using the Illumina platform. The *Arabidopsis thaliana* genome was used as a reference genome to align the reads and the MAQ software was used for sequence alignment. It was found that Illumina sequencing of the transcriptome was an efficient method for the identification of sequence variation between plant lines (Trick *et al.*, 2009). The *de novo* assembly of the transcriptome was found to be very effective when compared to the known *Arabidopsis thaliana* reference genome (Wellmer *et al.*, 2006).

The *Acropora millepora* coral transcriptome was subjected to 454 high-throughput pyrosequencing in order to provide a catalogue of expressed genes in the coral larvae (Meyer *et al.*, 2009). The study utilized publicly available software and basic tools requiring no previous sequence knowledge, a fact that shows merit in the sequencing of non-model organisms and organisms that have little to no genome sequence data available. A total of 146 Mb of data was generated with an average read length of 232 bases. Over 107 101 assembled sequences were generated from this data. Many sequences were then assigned names and functions according to the NCBI's database and gene ontology terms were assigned based on sequence similarity. The transcriptome coverage of the sequencing effort was evaluated by searching for genes involved in metabolic pathways and essential protein complexes. Failure to do so would mean

© University of Pretoria

incomplete sequencing. In both aspects close to 100% of the targets were found. Based on these comparisons, the completeness of the data was indirectly inferred.

The pyrosequencing platform was again used in a study of ESTs from a butterfly *Melitaea cinxia* and compared to traditional Sanger sequencing (Vera *et al.*, 2008). Complementary DNA was synthesized and ESTs were assembled with SEQMAN software. The transcriptome that was generated was then aligned to various other eukaryotes to estimate the coverage of the butterfly transcriptome. The study revealed that although a defined estimate of transcriptome coverage is unattainable without a full genomic sequence, most of the transcriptome may have been covered. The depth achieved by the 454 was much greater than that of Sanger sequencing, but if clones were not generated by the latter technology then the isolation of full length sequence data would have to rely on methods such as RACE. The depth achieved also allows higher sensitivity to SNP discovery but because the 454 sequences were derived from both strands, directional orientation could not be inferred. The microarray generated from this study demonstrated that 454 pyrosequencing was sufficient to characterize a draft transcriptome.

The 454 and Illumina sequencing platforms were used for the *de novo* assembly of *Pseudomonas syringae* pv. *oryzae* genome (Reinhardt *et al.*, 2009). These two platforms were used in concert to complement one another because both Illumina and 454 sequencing have drawbacks; 454 suffers from insertion and deletions and Illumina has difficulty in repeat assembly due to the short read lengths, but both these methods are much cheaper than Sanger sequencing. To validate the results, the already sequenced *P. syringae* pv. tomato DC300 reference genome was reassembled using the same HTS platforms. The results indicated that *de novo* assembly of a bacterial genome is achievable without a reference genome using HTS and by combining the two platforms, genome coverage was high and error rates were low (Reinhardt *et al.*, 2009). Studies using the 454 and Illumina platforms in conjunction with one

© University of Pretoria

another revealed that *de novo* assembly of microbial genomic data was greatly improved when using the two platforms over either one (Metzker, 2010).

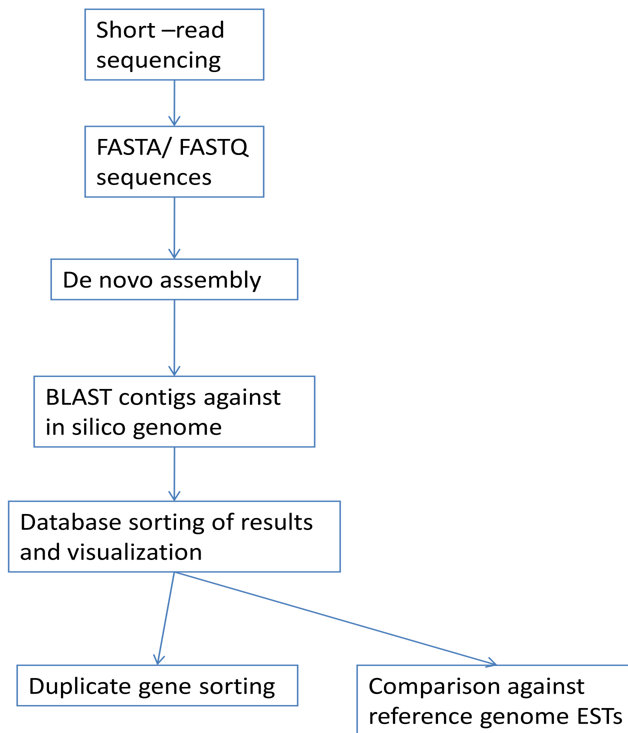


Figure 1. Overview of method used for transcriptome analysis using short-read sequences (Adapted from Collins *et al.,* 2008)**.**

Aside from using multiple sequencing strategies to obtain the best results there are other factors that influence the quality of the data obtained. Factors, such as random priming or oligo-dT priming for cDNA library generation and normalization of the cDNA library before sequencing (Wall *et al.*, 2009). A "simulation engine" called ESTstat was developed in order to determine the outcomes of different sequencing strategies using different technologies. The prediction is

© University of Pretoria

made using a variety of parameters (see Wall *et al.* 2009 for more details of the parameters) and testing had shown its efficacy even without a reference genome sequence. The software was then used to test the best combinations of sequencing technologies based on mean read length, transcriptome coverage and full length gene generation. The best results in the absence of a reference genome would be a combination of Illumina and 454 sequencing (Wall *et al.*, 2009).

Today's sequencing technologies are 200 times faster than those of the previous century and nearly 1000 times cheaper, this however is not the end. Faster, cheaper technologies are being produced and nanopore methods are being developed and are considered "next, next generation" sequencing technologies (Coombs, 2008). These cost effective future technologies will most probably rely on single molecule sequencing and allow unrivalled depth of sequencing, full transcriptome sequence generation or the generation of a large numbers of ESTs in smaller experiments (Simon *et al.*, 2009; Wall *et al.*, 2009).

The invention of high-throughput sequencing will influence and aid in current molecular tools such as phylogenetics, molecular fingerprinting, linkage mapping, and marker assisted selection. Another important application of high-throughput sequencing is its ability to allow the sequencing and cloning of single genes for their annotation and in order to determine the genes role in inherited diseases in e.g. humans. This technology further forms a cornerstone for genome annotation and hypothesis generation. The number of reads vs. the read length of sequence data generated is more important for gene discovery when the organism sequenced can be annotated from closely related sources (Priest *et al.*, 2010). However, read length outweighs read density when a reference genome's sequence data is not available or the organism in question is too distantly related to any model species to make valuable assumptions.

© University of Pretoria

1.4 Conclusion

Avocado production is worth nearly $7.5 billion in retail prices (Shumeta, 2010) but despite the importance of avocado and a 60 year attempt to unravel the host pathogen interaction, PRR is still the most limiting biotic factor in avocado production. Root pathogens such as *P. cinnamomi* and their interactions with hosts are poorly understood and there is a lack of molecular knowledge on avocado.

The deficiency of molecular information available for avocado is however supplemented by the ongoing work in marker development for the improvement of avocado production but there is still a need for much improvement (Chen *et al.*, 2007). The greater part of molecular detail exists due to a continuous effort in marker development to assist in either elucidating genetic relationships amongst scions (Acheampong *et al.*, 2008; Chanderbali *et al.*, 2008; Chang *et al.*, 1993; Clegg & Davis, 1989; Davis *et al.*, 1998; Mhameed *et al.*, 1997), or scion improvement (Chen *et al.*, 2007; Clegg *et al.*, 1992; Clegg *et al.*, 1999; Lavi *et al.*, 1991). Marker identification for tolerance to *P. cinnamomi* will result in the reduction of time required for rootstock breeding in the avocado industry. There is also a lack of understanding of the host-pathogen interactions of avocado and *P. cinnamomi.* The creation of genetic sequence resources will aid in unravelling the genetic basis of disease resistance and in the development of molecular markers that can be used for improved breeding and screening processes in the avocado industry (Acheampong *et al.*, 2008; Chanderbali *et al.*, 2008; Lavi *et al.*, 1991).

Genome sequence resources are fast being developed for economically important crops. The *Phytophthora* and plant research communities have united with a common cause to shed light on the defence mechanisms in plants (Huitema *et al.*, 2004). Currently there are many public databases available for *Phytophthora* species and their hosts, but these databases are lacking

© University of Pretoria

information regarding avocado. The data generated from a transcriptomic study of avocado in response to *P. cinnamomi* can contribute to this common cause.

A rapid, cost effective method to create genetic data for non-model crops such as avocado is high-throughput sequencing. The technology can be used to address the lack of molecular information available for avocado and to gain an understanding of the avocado- *P. cinnamomi* interaction. After having compared the three current platforms for high-throughput sequencing it is apparent that the lack of a reference genome to align sequence reads to will be a limitation using small read lengths of the Illumina and SOLiD platforms (Holt & Jones, 2008; Priest *et al.*, 2010). The short length of the sequence reads for both platforms along with their difficulty in assembly will not be suitable for *de novo* sequencing of a non-model crop such as avocado, and we thus conclude that the 454 platform by Roche will be the most suitable platform.

The generation of molecular information on the avocado- *P. cinnamomi* interaction will allow research questions such as: which genes are transcribed, what are the expression profiles of defence-related genes and at what point does expression patterns have a maximal increase to be answered. Due to the importance of avocado as a fruit crop and the insight that the basal angiosperm provides to other magnolids or phylogenetically related plants, it is important that we expand on the knowledge base of molecular information. Detailed molecular studies will allow avocado to become a model for basal angiosperms in functional genomics - a group which so far has no other model (Chanderbali *et al.*, 2008). The generation of sequence data and molecular information for the avocado will lead to more detailed studies to be conducted and supplement studies such as those by García-Pineda *et al.* (2010) and Rookes *et al.* (2008) in plant pathogen interaction.

© University of Pretoria

1.5 References

Acheampong K. A., Akromah R., Ofori F. A., Takarama J. F., Saada D., Bitton I., Lavi U. (2008) Genetic characterization of Ghanaian avocados. *Journal of the American Society for Horticultural Science* **133**, 801-809.

Ahmadian A., Ehn M., Hober S. (2006) Pyrosequencing: history, biochemistry and future. *Clinica Chimica Acta* **363**, 83-94.

Alcaraz M. L., Hormaza J. I. (2007) Molecular characterization and genetic diversity in an avocado collection of cultivars and local Spanish genotypes using SSRs. *Hereditas* **144**, 244-253.

Ashworth V. E. T. M., Clegg M. T. (2003) Microsatellite markers in avocado (*Persea americana* Mill.): Genealogical relationships among cultivated avocado genotypes. *Journal of Heredity* **94**, 407-415.

Ashworth V. E. T. M., Kobayashi M. C., De La Cruz M., Clegg M. T. (2004) Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. *Scientia Horticulturae* **101**, 255-267.

Ben-Ya'acov A., Michelson E. (1995) Avocado rootstocks. *Horticultural Reviews* **17**, 381-429.

Bergh B., Ellstrand N. (1986) Taxonomy of the avocado. *California Avocado Society Yearbook* **70**, 135-145.

Bijzet Z. (1999) Developing and optimizing techniques to facilitate the development of avocado rootstocks resistant to *Phytophthora cinnamomi*. *South African Avocado Growers' Association Yearbook* **22**, 96-100.

© University of Pretoria

Borrone J. W., Schnell R. J., Violi H. A., Ploetz R. C. (2007) Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. *Molecular Ecology Notes* **7**, 439-444.

Cass L. G., Kirven K. A., Christoffersen R. E. (1990) Isolation and characterization of a cellulase gene family member expressed during avocado fruit ripening. *Molecular and General Genetics* **223**, 76-86.

Chanderbali A. S., Albert V. A., Ashworth V. E. T. M., Clegg M. T., Litz R. E., Soltis D. E., Soltis P. S. (2008) *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. *BioEssays* **30**, 386-396.

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

Chen H., Ashworth V. E. T. M., Xu S., Clegg M. T. (2007) Quantitative genetic analysis of growth rate in avocado. *Journal of the American Society for Horticultural Science* **132**, 691-696.

Chernys J. T., Zeevaart J. A. D. (2000) Characterization of the 9-cis-epoxycarotenoid dioxygenase gene family and the regulation of abscisic acid biosynthesis in avocado. *Plant Physiology* **124**, 343-354.

Christoffersen R. E., Tucker M. L., Laties G. G. (1984) Cellulase gene expression in ripening avocado fruit: The accumulation of cellulase mRNA and protein as demonstrated by cDNA hybridization and immunodetection. *Plant Molecular Biology* **3**, 385-391.

Christoffersen R. E., Warm E., Laties G. G. (1982) Gene expression during fruit ripening in avocado. *Planta* **155**, 52-57.

Clegg M. T., Davis J. W. (1989) Molecular genetics of avocado. *California Avocado Society Yearbook*, 59-61.

© University of Pretoria

Clegg M. T., Henderson D., Durbin M. (1992) The development of molecular markers in avocado. *Proceedings of Second World Avocado Congress 1992, 573.*

Clegg M. T., Kobayashi M., Lin J. Z. (1999) The use of molecular markers in the managment and improvement of avocado (*Persea americana* Mill.). *Revista Chapingo Serie Horticultura* **5**, 227-231.

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

Coffey D. M., Guillemet F. B. (1987) Avocado rootstocks. *California Avocado Society Yearbook* **71**, 173-179.

Collard B. C. Y., Jahufer M. Z. Z., Brouwer J. B., Pang E. C. K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**, 169-196.

Collins L. J., Biggs P. J., Voelckel C., Joly S. (2008) An approach to transcriptome analysis of non-model organisms using short read sequences. *Genome Informatics* **21**, 3-14.

Coombs A. (2008) The sequencing shakeup. *Nature Biotechnology* **26**, 1109-1112.

Davis J., Henderson D., Kobayashi M., Clegg M. T. (1998) Genealogical relationships among cultivated avocado as revealed through RFLP analyses. *Journal of Heredity* **89**, 319-323.

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

Durban J., Juarez P., Angulo Y., Lomonte B., Flores-Diaz M., Alape-Giron A., Sasa M., Sanz L., Gutierrez J., Dopazo J., Conesa A., Calvete J. (2011) Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. *BMC Genomics* **12**, 259.

© University of Pretoria

Elam P. (1997) Buding and grafting citrus and avocados in the home garden. In: *Home orchad*. University of California **8001**.

Faber B. A., Morse J. G., Hoddle M. S., Eskalen A., Bender G. S., Lanini W. T., Wilen C. A., Phillips P. A., Shrestha A. (2010) *The UC IPM Pest Management Guidelines: Avocado* UC Statewide IPM Program, Davis.

Farrer R. A., Kemen E., Jones J. D. G., Studholme D. J. (2009) *De novo* assembly of the *Pseudomonas syringae* pv. *syringae* B728a genome using Illumina/Solexa short sequence reads. *Federation of European Microbiological Societies Microbiology Letters* **291**, 103-111.

Fernandes J., Morrow D. J., Casati P., V. W. (2008) Distinctive transcriptome responses to adverse environmental conditions in *Zea mays* L. *Plant Biotechnology Journal* **6**, 782-798.

Fiedler J., Bufler G., Bangerth F. (1998) Genetic relationships of avocado (*Persea americana* Mill.) using RAPD markers. *Euphytica* **101**, 249-255.

Gabor B. K., Coffey D. M. (1991) Comparison of rapid methods for evaluating resistance to *Phytophthora cinnamomi* in avocado rootstocks. *Plant Disease* **75**, 118-120.

Gabor B. K., Guillemet F. B., Coffey D. M. (1990) Comparison of field resistance to *Phytophthora cinnamomi* in twelve avocado rootstocks. *Hortscience* **25**, 1655-1656.

García-Pineda E., Benezer-Benezer M., Gutiérrez-Segundo A., Rangel-Sánchez G., Arreola-Cortés A., Castro-Mercado E. (2010) Regulation of defence responses in avocado roots infected with *Phytophthora cinnamomi* (Rands). *Plant and Soil* **331**, 45-56.

Giblin F., Pegg K., Willingham S., Anderson J., Coates L., Cooke T., Dean J., Smith L. (2005) *Phytophthora* Revisited. *New Zealand and Australia Avocado Grower's Conference*, Tauranga, 1-7.

© University of Pretoria

Govers F., Gijzin M. (2006) *Phytophthora* genomics: The plant destroyers' genome decoded. *The American Phytopathological Society* **19**, 1295-1301.

Hinch J. M., Clarke A. E. (1982) Callose formation in *Zea mays* as a response to infection with *Phytophthora cinnamomi*. *Physiological Plant Pathology* **21**, 113-124.

Holt R. A., Jones S. J. M. (2008) The new paradigm of flow cell sequencing. *Genome Research* **18**, 839-846.

Huitema E., Bos J. I. B., Tian M., Win J., Waugh M. E., Kamoun S. (2004) Linking sequence to phenotype in *Phytophthora*-plant interactions. *Trends in Microbiology* **12**, 193-200.

Jackson B., Schnable P., Aluru S. (2009) Parallel short sequence assembly of transcriptomes. *BMC Bioinformatics* **10**, 14.

Johnson D. S., Mortazavi A., Myers R. M., Wold B. (2007) Genome-wide mapping of *in-vivo* protein-DNA interactions. *Science* **316**, 1497-1502.

Kaiser C., Whiley A. W., Pegg K. G., Hargreaves P. A., Weinhart M. P. (1997) Determination of critical root concentrations of phosphate to control Phytophthora root rot in avocado. *Joint meeting of the Australian Avocado Grower's Federation Inc. and the New Zealand Avocado Grower's Association*, 155.

Keane P. J. (1981) *Phytophthora cinnamomi* and the diseases it causes. *Soil Biology and Biochemistry* **13**, 559-559.

Kellam M. K., Coffey D. M. (1985) Quantitative comparison of the resistance to Phytophthora root rot in three avocado rootstocks. *Phytopathology* **75**, 230-234.

Kobayashi M., Lin J.-Z., Davis J., Francis L., Clegg M. T. (2000) Quantitative analysis of avocado outcrossing and yield in California using RAPD markers. *Scientia Horticulturae* **86**, 135-149.

© University of Pretoria

Köhne J. S. (2005) Selection of Avocado Scions and breeding of rootstocks in South Africa. *New Zealand and Australia Avocado Grower's Conference*, Tauranga, New Zealand.

Kremer-Köhne S., Mukhumo M. L. (2003) Breeding and field evaluation of new rootstocks for increased Hass yields and resistance to root rot in South Africa. *Proceedings of V World Avocado Congress*, 555-560.

Kumar L. S. (1999) DNA markers in plant improvement: An overview. *Biotechnology Advances* **17**, 143-182.

Lavi U., Akkaya M., Bhagwat A., Lahav E., Cregan P. B. (1994) Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana* M.). *Euphytica* **80**, 171-177.

Lavi U., Hillel J., Vainstein A., Lahav E., Sharon D. (1991) Application of DNA fingerprints for identification and genetic analysis of avocado. *Journal of the American Society for Horticultural Science* **116**, 1078-1081.

Lavi U., Sharon D., Mhameed S., Kashkush H., Adato A., Kaufman D., Cregan P. B., Lahav E., Tomer E., Hillel J. (1998) Molecular markers in tropical and subtropical horticulture. *International symposium on Biotechnology of tropical and subtropical species* (ed. Drew R. A.), 49-54. International Society for Horticultural Science.

Li P., Peatman E., Wang S., Feng J., He C., Baoprasertkul P., Xu P., Kucuktas H., Nandi S., Somridhivej B., Serapion J., Simmons M., Turan C., Liu L., Muir W., Dunham R., Brady Y., Grizzle J., Liu Z. (2007) Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs. *BMC Genomics* **8**, 177.

Linde C. (1999) *Population structure of Phytophthora cinnamomi in South Africa* PhD Dissertation, University of the Orange Free State.

Mardis E. R. (2007) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**, 133-141.

© University of Pretoria

Margulies M., Egholm M., Altman W. E., Attiya S., Bader J. S., Bemben L. A., Berka J., Braverman M. S., Yi-Ju C., Zhoutao C., Dewell S. B., Lei D., Fierro J. M., Gomes X. V., Godwin B. C., Wen H., Helgesen S., Chun He H., Irzyk G. P., Jando S. C. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-380.

McGarvey D. J., Sirevåg R., Christoffersen R. E. (1992) Ripening-related gene from avocado fruit: Ethylene-inducible expression of the mRNA and polypeptide. *Plant Physiology* **98**, 554-559.

Menge J. A. (1999) Screening and evaluation of new rootstocks with resistance to *Phytophthora cinnamomi*. *California Avocado Research Symposium*, 69-72. California Avocado Society, University of California, Riverside, USA.

Metzker M. L. (2010) Sequencing technologies- the next generation. *Nature Review Genetics* **11**, 31-46.

Meyer E., Aglyamova G., Wang S., Buchanan-Carter J., Abrego D., Colbourne J., Willis B., Matz M. (2009) Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFLX. *BMC Genomics* **10**, 219.

Mhameed S., Sharon D. (1996) Level of heterozygosity and mode of inheritance of variable number of tandem repeat loci in avocado. *Journal of the American Society for Horticultural Science* **121**, 778-782.

Mhameed S., Sharon D., Kaufman D., Lahav E., Hillel J., Degani C., Lavi U. (1997) Genetic relationships within avocado (*Persea americana* Mill) cultivars and between *Persea* species. *Theoretical and Applied Genetics* **94**, 279-286.

Morozova O., Marra M. A. (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255-264.

© University of Pretoria

Priest H. D., Fox S. E., Filichkin S. A., Mockler T. C. (2010) Utility of next-generation sequencing for analysis of horticultural crop transcriptomes. *International Symposium on Molecular Markers in Horticulture*, 283-288. International Society for Horticultural Science Acta Horticulturae.

Reinhardt J. A., Baltrus D. A., Nishimura M. T., Jeck W. R., Jones C. D., Dangl J. L. (2009) *De novo* assembly using low-coverage short read sequence data from the rice pathogen *Pseudomonas syringae* pv. *oryzae*. *Genome Research* **19**, 294-305.

Ronaghi M., Uhlen M., Nyren P. (1998) A sequencing method based on real-time pyrophosphate. *Science* **281**, 363-365.

Rookes J. E., Wright M. L., Cahill D. M. (2008) Elucidation of defence responses and signalling pathways induced in *Arabidopsis thaliana* following challenge with *Phytophthora cinnamomi*. *Physiological and Molecular Plant Pathology* **72**, 151-161.

Rothberg J. M., Leamon J. H. (2008) The development and impact of 454 sequencing. *Nature Biotechnology* **26**, 1117-1124.

Scheibye-Alsing K., Hoffmann S., Frankel A., Jensen P., Stadler P. F., Mang Y., Tommerup N., Gilchrist M. J., Nygård A. B., Cirera S., Jørgensen C. B., Fredholm M., Gorodkin J. (2009) Sequence assembly. *Computational Biology and Chemistry* **33**, 121-136.

Schnell R. J., Brown J. S., Olano C. T., Power E. J., Krol C. A. (2003) Evaluation of avocado germplasm using microsatellite markers. *Journal of the American Society for Horticultural Science* **128**, 881-889.

Sharon D., Cregan P. B., Mhameed S., Kusharska M., Hillel J., Lahav E., Lavi U. (1997) An integrated genetic linkage map of avocado. *Theoretical and Applied Genetics* **95**, 911-921.

Sharon D., Hillel J., Mhameed S., Cregan P. B., Lavi U., Lahav E. (1998) Association between DNA markers and loci controlling avocado traits. *American Society for Horticultural Science* **123**, 1016-1022.

© University of Pretoria

Shumeta Z. (2010) Avocado production and marketing in Southwestern Ethiopia. *Trends in Agricultural Economics* **3**, 190-206.

Simon S. A., Zhai J., Nandety R. S., McCormick K. P., Zeng J., Mejia D., Meyers B. C. (2009) Short-read sequencing technologies for transcriptional analyses. *Annual Review of Plant Biology* **60**, 305-333.

Sorber K., Chiu C., Webster D., Dimon M., Ruby J. G., Hekele A., DeRisi J. L. (2008) The long march: a sample preparation technique that enhances contig length and coverage by high-throughput short-read sequencing. *PLoS ONE* **3**, e3495.

Torres A. M., Bergh B. O. (1980) Fruti and leaf isozymes as genetic markers in avocado. *Journal of the American Society for Horticultural Science* **105**, 614-619.

Torres A. M., Diedenhofen U., Bergh B. O., Knight R. J. (1978) Enzyme polymorphisms as genetic markers in the avocado. *American Journal of Botany* **65**, 134-139.

Traub, Hamilton P., Ralph Robinson T. (1937) Avocado. *Improvement of Subtropical Fruit other than Citrus*, 10-26. United States Department of Agriculture Yearbook.

Trick M., Long Y., Meng J., Bancroft I. (2009) Single nucleotide polymorphism (SNP) discovery in the polyploid *Brassica napus* using solexa transcriptome sequencing. *Plant Biotechnology Journal* **7**, 1-13.

Varshney R. K., Graner A., Sorrells M. E. (2005) Genomics-assisted breeding for crop improvement. *Trends in Plant Science* **10**, 621-630.

Vera J. C., Wheat C. W., Fescemyer H. W., Frilander M. J., Crawford D. L., Hanski I., Marden J. H. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**, 1636-1647.

© University of Pretoria

Wall P. K., Leebens-Mack J., Chanderbali A., Barakat A., Wolcott E., Liang H., Landherr L., Tomsho L., Hu Y., Carlson J., Ma H., Schuster S., Soltis D., Soltis P., Altman N., dePamphilis C. (2009) Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10**, 347.

Weber A. P., Weber K. L., Carr K., Wilkerson C., Ohlrogge J. B. (2007) Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiology* **144**, 32 - 42.

Wellmer F., Alves-Ferreira M. r., Dubois A., Riechmann J. L., Meyerowitz E. M. (2006) Genome-wide analysis of gene expression during early *Arabidopsis* Flower Development. *PLoS Genetics* **2**, e117.

Whiley A. W., Kohne J. S., Arpaia M. L., Bender G. S. (1990) Future prospects with new avocado cultivars and elite rootstocks. *South African Avocado Growers' Association Yearbook* **13**, 13-20.

Whiley A. W., Schaffer B., Wolstenholme B. N. (2002) *The Avocado – botany, production and uses* CABI Publishing, New York.

Wolstenholme B. N. (2003) Avocado rootstocks: What do we know; are we doing enough research? *South African Avocado Growers' Association Yearbook* **26**, 106-112.

Zentmyer G. A. (1984) Avocado diseases. *Tropical Pest Managment* **30**, 388-400.

Zentmyer G. A. (1955) Diseases of the Avocado. *California Avocado Society Yearbook* **39**, 44-58.

Zentmyer G. A. (1957) The search for resistant rootstocks in Latin America. *California Avocado Society Yearbook* **41**, 101-106.

© University of Pretoria

Zentmyer G. A., Thorn W. A. (1956) Resistance of the Duke variety of avocado to Phytophthora root rot. *California Avocado Society Yearbook* **40**, 169-173.

© University of Pretoria

CHAPTER 2

EST Sequencing of the avocado transcriptome in response to *Phytophthora cinnamomi*

© University of Pretoria

2.1 Introduction

Avocados are grown in more than 50 countries and its production is envisioned to reach an estimate of about $7.5 billion in retail prices (Shumeta, 2010). The crop consists of various cultivars with the most popular cultivar being Hass. Avocado production is greatly affected by Phytophthora root rot (PRR) caused by the root pathogen *Phytophthora cinnamomi* Rands (Coffey, 1987; Zentmyer, 1984; Zentmyer, 1955). Although being identified as the causal agent in 1929 *P. cinnamomi (Pc)* is still currently regarded as avocado's most devastating pathogen.

Methods of *Pc* control are under threat of becoming insufficient. Currently *Pc* is controlled by using clean planting material and high organic matter soils along with phosphite trunk injections and foliar spraying and finally with the use of tolerant rootstocks (Coffey, 1987; Giblin *et al.*, 2005; Kaiser *et al.*, 1997). Rootstock breeding has received attention since the commercialization of avocados in the 1900s (Ben-Ya'acov & Michelson, 1995) with resistance in rootstocks being discovered decades ago (Kellam & Coffey, 1985). Although the selection of disease tolerant rootstocks has become routine in the avocado industry, the process is arduous and may take up to 20 years to release a new cultivar (Köhne, 2005). Recently decreased sensitivity to the use of phosphite has been identified, and it is feared that completely resistant pathogen strains may arise (Dobrowolski *et al.*, 2008).

The search for genes offering resistance against pathogens is an important aspect in the fight against *Pc* (García-Pineda *et al.*, 2010; Rookes *et al.*, 2008) but molecular work on avocado is limited. Molecular work in avocado began with marker development (isozymes) in 1989 (Clegg & Davis, 1989). These isozyme markers were then succeeded by higher resolution restriction fragment length polymorphisms (RFLPs) (Clegg *et al.*, 1992). Other markers have been developed with the progression of time and technology, these include microsatellite markers

© University of Pretoria

and SSRs (simple sequence repeats) (Lavi *et al.*, 1994; Lavi *et al.*, 1991). These have been used to construct a genetic linkage map (Sharon *et al.*, 1997) in the hope of aiding in QTL (quantitative trait loci) identification for the selection and breeding of crops with desirable traits. Large genetic diversity within avocado has complicated marker development (Ashworth *et al.*, 2004) and although some progress has been made in marker identification, most studies to date have focused on fruiting and flowering characteristics (Chen *et al.*, 2007). An example of such a study is the use of RAPD (Random Amplification of Polymorphic DNA) and SSR markers to screen for QTLs associated with eight avocado fruit traits (fruit weight, skin colour, skin gloss, skin thickness, skin peeling, seed size, flesh fibres and taste) (Sharon *et al.*, 1998). In another study it was shown that RFLP markers were able to identify different rootstock sources and could potentially be used to select rootstocks with different levels of tolerance to *Pc* (Clegg *et al.*, 1999). Currently the most recent development in developing markers against *Pc* saw the generation of 70 microsatellite markers but are yet to be tested (Borrone *et al.*, 2007). The University of California Riverside (UCR) has recently employed 61 polymorphic AFLP markers to characterise PRR tolerance in 83 rootstocks from various locations including South Africa and Israel with the majority of rootstocks from the UCR collection (Douhan *et al.*, 2011). The study concluded that resistance mechanisms vary between tolerant cultivars and no trend was observed in the cluster analysis.

In avocado the genes expressed in response to pathogen infection are unknown and their study is an important step in understanding plant defence. There are various techniques available for analyzing genes differentially expressed and their associated patterns of expression. SSH (Diatchenko *et al.*, 1996), SAGE (serial analysis of gene expression) (Velculescu *et al.*, 1995), qRT-PCR (Gachon *et al.*, 2004), cDNA microarray analysis (Schummer *et al.*, 1997) and

© University of Pretoria

recently high-throughput sequencing platforms(Vega-Arreguin *et al.*, 2009; Wall *et al.*, 2009) have been used for gene discovery and expression analysis.

These platforms are efficient at generating large amounts of EST data (Scheibye-Alsing *et al.*, 2009; Vega-Arreguin *et al.*, 2009) and provide a tool for gene discovery in non-model crops (Priest *et al.*, 2010; Weber *et al.*, 2007). This allows better understanding of the changes in expression in response to abiotic or biotic stresses (Vera *et al.*, 2008; Weber *et al.*, 2007). The elimination of bacterial cloning (Simon *et al.*, 2009) and the lower cost involved in sequencing when compared to that of capillary sequencing (Wall *et al.*, 2009) makes HTS platforms ideal for mRNA analysis. Additionally, biases of instability in the cloning process are avoided with pyrosequencing of cDNA transcripts (Weber *et al.*, 2007). There are many types of high-throughput sequencing platforms available but the best for use with non-model organisms is pyrosequencing (Vera *et al.*, 2008). More specifically, it is advantageous for commercial crops that lack substantial molecular databases to aid in their improvement (Priest *et al.*, 2010). The pyrosequencing platform is advantageous due to the lack of a reference genome, which is a limitation in using RNA-seq from Illumina (Priest *et al.*, 2010).

Nothing is known regarding the molecular process involved in resistance or metabolic pathways of avocado in the avocado-*Pc* interaction. Due to the pressures of *P. cinnamomi* on avocado production a better understanding of molecular processes involved in resistance will aid in the arms race against this devastating root pathogen. The application of HTS in avocado research is the next step in aiding in the improvement of this economically important crop. The aim of this study was to generate sequence data for gene mining to identify genes involved in the tolerant avocado-*Pc* interaction.

© University of Pretoria

2.2 Materials & Method

2.2.1 Avocado inoculation with *P. cinnamomi* &harvesting of root material

Nine-month-old tolerant R0.09 clonal avocado (*Persea americana*) plantlets were provided by Westfalia Technological Services (Tzaneen, South Africa) and inoculated with *Pc* mycelia. A total of 33 g of mycelia was homogenised in 65 L of distilled water resulting in a final concentration of 0.5 g/L. This was then mixed into 112 kg of vermiculite in a mistbed. Plantlets were randomly grounded in vermiculite and constantly irrigated over a period of six weeks. Root material was harvested at 0 hour (uninfected), 3, 6, 12, 24, 48 and 72 hours post infection (hpi), snap frozen in liquid nitrogen and stored on dry ice (-78 °C) until the root material could be placed in a -80 °C freezer. A subset of plants was left in the mistbed for six weeks after which disease symptoms were recorded. The pathogen was re-isolated and the true identity was confirmed using *Pc* specific LPV3 forward (5'-GTGCAGACTGTCGATGTG-3') and LPV3 reverse (5'-GAACCACAACAGGCACGT-3') primers that amplify a 450 bp *Pc* specific fragment (Kong *et al*., 2003.).

2.2.2 RNA extraction

RNA isolations were done using the CTAB method (Chang *et al.*, 1993). Roots were ground in liquid nitrogen and 2-3 grams of ground root material were used per RNA extraction. The Chloroform: isoamyl alcohol wash step was repeated 6 times followed by washing with ethanol thrice. Total RNA concentrations were quantified using the Nanodrop ND-100 Spectrophotometer (Nanodrop Technologies, Inc., Montchanin, USA) and verified on a 2% non-

© University of Pretoria

denaturing TAE agrose gel. Total RNA from three biological replicates (per time point) was combined before mRNA purification.

2.2.3 mRNA isolation & cDNA synthesis:

Prior to mRNA isolation, total RNA time points were combined into three libraries for pyrosequencing. The 0 hour time point was regarded as the uninfected library, infection library 1 was composed of 6, and 12 hours, while infection library 2 comprised of 24, 48 and 72 hour samples (the 3 hour time point was not used for pyrosequencing but for the expression analysis of chapter 3). Purification of mRNA was done on equal amounts of RNA (100 µg) according to manufacturer's instructions using Oligotex (Oligotex™ mRNA kit, Qiagen, Valencia, California, USA). The mRNA purification was performed twice per sample to ensure the removal of any DNA contamination as well as to reduce the amount of rRNA available in the sample.

The mRNA isolation was tested for DNA contamination using flavanone-3-hydroxylase (F3H) primers from Inqaba Biotec (Sunnyside, South Africa). F3H forward (5'-TCTGATTTCGGAGATGACTCGC-3') and F3H reverse (5'- TGTAGACTTGGGCCACCTCTTT-3') primers amplify a 300 bp fragment of the flavanone-3-hydroxylase gene from RNA as opposed to the 1200 bp fragment which is obtained from DNA. After isolation of mRNA from total RNA, RiboLock RNase inhibitor (Fermentas, Life Sciences, Hanover, USA) was added to each sample and analyzed by gel electrophoresis to monitor degradation.

A standard PCR protocol was used for the testing of DNA contamination. Amplification reactions composed of 2 µl of reaction buffer, 1.5 µl MgCl, 1 µl DNA template, 11.3 µl H_2O, 10 mM F3H forward primer, 10 mM F3H reverse primer, 2 mM dNTP's and 0.2 µl (1 U /µl) Taq polymerase were used in a single reaction volume of 20 µl. All reagents for amplification were from Bioline

© University of Pretoria

(Bioline, Taunton, USA). Amplification cycles were as follows: 94ºC for 2 minutes, then 30 cycles of 94ºC for 30 seconds, 60ºC for 30 seconds and 72ºC for 30 seconds followed by final elongation of 72ºC for 7 minutes.

The cDNA libraries were synthesized with the Roche cDNA synthesis system (Roche Diagnostics, Mannheim, Germany) according to manufacturer's instructions. Three cDNA libraries were generated for the various infection stages mentioned before, namely uninfected, infection library 1 and infection library 2. First strand synthesis was performed by adding 2 µl of oligo dT_{15} primer and 19 µl of mRNA to obtain a total volume of 21 µl. Incubation at 70ºC for 10 min allowed binding of the primer and the reaction was terminated by placing the sample on ice. The following was then added: 8 µl Reverse Transcriptase (RT) buffer, 4 µl 0.1 M dithiothreitol (DTT), 2 µl Avian Myeloblastosis Virus (AMV), 1 µl RNase inhibitor (25U/µl) and 4 µl 10 mM dNTP-mix were added to the first strand reaction. The reaction was then incubated at 42ºC for 60 min and placed on ice to terminate the reaction. Second strand synthesis was continued directly after first strand synthesis by adding 40 µl of first strand cDNA, 30 µl of 2nd strand synthesis buffer, 1.5 µl dNTP-mix, 6.5 µl 2nd strand enzyme blend and 72 µl water. The total reaction (150 µl) was incubated for 2 hours at 16ºC and then terminated by adding 17 µl EDTA 0.2M pH 8.

Complementary DNA was purified using the Qiagen MinElute PCR Purification Kit (Qiagen, Valencia, California, USA) before being sequenced. The total volume of cDNA was ~180 µl (approximately 25 ng/µl) so the reaction was split into two separate volumes of 90 µl each, a total of 300 µl ERC buffer was added to the cDNA solutions before application to the spin columns. Samples were centrifuged at 10 000 rpm for 1 min to bind the cDNA to the column. The sample was then washed with 750 µl PE buffer and the centrifugation step was repeated twice to remove all traces of ethanol.

© University of Pretoria

The cDNA was eluted in sterile water with four different treatments being tested (a single elution using standard volume as indicated in the instructions, using a slightly higher volume to elute cDNA in, a double elution through the same spin column and a dual elution through separate spin columns) to try and increase the amount of eluted cDNA. Analysis of the cDNA cleanup was done by gel electrophoresis on a 1.5% TAE agrose gel run at 80 V for ~30 min. The approximate size of the cDNA fragments ranged from 3 kb to 400 bp (which is desired).

DNA contamination was assessed using flavanone-3-hydroxylase (F3H) primers as described previously in paragraph 3 on page 59.

2.2.4 Pyrosequencing & Data annotation

The three cDNA libraries were sequenced by Inqaba Biotec (Sunnyside, South Africa) on a single lane of the 454 GS-FLX platform. To facilitate the sequencing three different nucleotide tags were attached to each sample. The adapter contains a recognition sequence which is used to identify the sample and is removed during the bioinformatic analysis.

Contigs were assembled using the Newbler assembler version 1.1.02.15 (Roche) by bioinformaticists at Inqaba Biotec. These contigs were then re-analyzed and annotated using dCAS (Desktop cDNA Annotation System) Version 1.4.1 Build 3791 and CLC Main Workbench software (CLC bio, Cambridge, MA) by myself. Reads were trimmed before contig assembly with low quality reads not being included in the assembly. The contigs were also ranked against the Gene Ontology (GO) and EuKaryotic Orthologous Groups (KOG) databases. This gave insight into the functional classes of the sequence data generated.

The BLASTX tool was used to search for protein homologies on the NCBI database (Altschul *et al.*, 1990). This was done using the PAM 30 matrix searching for short and nearly exact

© University of Pretoria

matches. Comparing amino acid sequences is a more reliable predictor of similarity between two sequences than comparing nucleotide sequences because of the considerable variation in nucleotide sequences and the evolutionary distance of the avocado. Contigs were selected based on their role in defence and further identified by identifying the conserved domains using the conserved domain database. The top 5 BLASTX homologous sequences from the NCBI were used to perform alignments of selected defence-related genes using CLC software.

2.3 Results & Discussion

2.3.1 Avocado inoculation with *P. cinnamomi*

Inoculation of the avocado plantlets resulted in the development of PRR symptoms within six weeks. The plantlets showed signs of wilting and deterioration in overall plant health. Upon inspection of the roots it was found that vascular discolouration had begun and the young feeder roots that had been infected showed signs of necrosis. *Phytophthora cinnamomi* was re-isolated from diseased roots, proving Koch's postulates. The *P. cinnamomi* specific LPV amplification generated bands of ~450 bp from the re-isolated material indicating presence of the pathogen. These results indicate that the infection method was successful and caused disease.

© University of Pretoria

2.3.2 RNA extraction

RNA extraction yield was very high with values averaging on 1736 ng/ µl from the harvested roots. The purity of the RNA was indicated by a 260/280 ratio of ~2.11 and a 260/230 ratio of ~2.5. These values measure protein and organic solvents respectively, with a value of ~2 indicating good quality RNA. Distinct 28S and 18S bands were visualised for all RNA extractions with no severe degradation being visualized (Figure 1).

2.3.3 mRNA isolation & cDNA synthesis:

Total RNA from three biological replicates per time point (100 µg) were pooled into three libraries for mRNA isolation (uninfected, infection libraries 1 and 2). Successful cDNA synthesis resulted in a smear between 500 bp and 3 kb (Figure 2a).

Amplification using the F3H primers indicated that there was no DNA contamination yielding the expected amplicon of ~240 bp for cDNA compared to the 1200 bp fragment expected for genomic DNA. The difference in band size is due to the removal of the intron sequence in cDNA. All samples were negative for the 1200 bp band but positive for the smaller cDNA band (Figure 2b).

2.3.4 Pyrosequencing & data annotation

Pyrosequencing of the three libraries (uninfected, infection libraries 1 and 2) were accomplished in a single lane using three tags (Table 1) and resulted in a total of 2 Mb (of 2.5 Mb) of data being generated (after trimming and quality control). The sequencing depth of this study is not

© University of Pretoria

sufficient to make direct conclusions about differential gene expression and shall be supplemented with expression analysis of selected defence-related genes in chapter 3.

A total number of 9953 reads were generated comprising of 1407 reads from the uninfected, 3584 reads from infection library 1 and 4962 reads from the infection library 2. Average read lengths for the libraries were; 216.4 for uninfected, 217.5 and 215.9 for infection libraries 1 and 2 respectively, which generated an average contig length of 250 bp. 12. The average contig length is very similar to the read length because the parameters for contig length were set to be low so as to characterise as many transcripts as possible as opposed using large contig length parameters to identify full length transcripts. Contigs comprised of approximately 5 reads with a 2 read minimum. In total 371 contigs were generated (Table 2 & Figure 3). The differences in the total amount of data generated per library could possibly be attributed to the differences in cDNA concentrations combined with only a third of the sample being used for pyrosequencing of each library.

Three hundred and sixty seven novel ESTs were identified for *P. americana* using the dCAS. The program used BLASTX amino acid comparisons to screen the contigs against the NCBI non-redundant (NR) database. Of the 371 contigs identified, only four showed homology to avocado EST's on the database. Three of these transcripts encoded for the metallothionein type II protein and one for fructose- bisphosphate aldolase with the remaining 367 having not been previously identified. Contigs that were not ascribed any description according to the automatic BLAST were then annotated manually. The manual BLAST did not yield any significant data for those reads. A total of 46% of transcripts showed homology to specific genes while 41% were assigned to hypothetical proteins. Only 12% of transcripts could not be identified (Table 3).

Contigs were further grouped into functional classes according to the GO (Gene Ontology) and KOG (Eukaryotic Orthologous Groups) databases. Nine percent of contigs had unknown

© University of Pretoria

function in the KOG database while 44.5 % of contigs from the GO classification were represented by unknown functions (Figure 4 & Table 7). According to the KOG database 34.8% of contigs were represented in the categories post-translational modification; translation, ribosomal structure and biogenesis; signal transduction mechanisms and general function. The GO database classified 40.8% of contigs as being involved in the cell wall; protein binding; stress response; ribosomal structural; cytoplasmic biological processes and cellular components. Only 3% of the contigs were directly associated with stress responses. Over 20 putative defence related genes were identified although this is not reflected in the GO or KOG classifications due to their classification into other processes. Defence-related genes included genes involved in general defence (metallothioneins, thaumatin and universal stress proteins) to genes more specific to oomycete defence (pathogenesis related protein PR10, oxysterol binding protein) (Table 3). The number of defence-related genes identified is relatively small and is expected to be larger if more sequencing is done. I would however expect that the number of defence-related genes identified in the infected libraries vs. that of the uninfected library is significant and more than what would be obtained from randomly sequencing a healthy plant.

2.3.4.1 EST identification & classifications

A total of 371 contigs were generated from the three cDNA libraries. Of the 371 contigs, 192 contigs had positive gene identities, while 123 contigs were identified only as hypothetical proteins. There were 43 contigs that could not be aligned to any sequence in the NR database based on nucleotide or amino acid sequence. Contigs that had positive plant homologies supported with strong E-values were then further characterized by assigning identities according to the NR database of the NCBI. ESTs are discussed under their respective cDNA libraries in

© University of Pretoria

three super classifications (primary metabolism, stress response and cell wall associated) (Table 3).

As expected, transcripts identified in the uninfected avocado library were mainly involved in metabolism, transcription and RNA processing. The uninfected library contained 43 contigs with 23 gene transcripts showing homology to sequences in the NCBI NR database. Seventeen were homologous to plants and six showed homology out of the plant kingdom. A further 14 contigs were ascribed only as hypothetical proteins and one contig as an unknown protein while five contigs showed no homology to any sequence on the database (Table 4).

Infected library 1 contained 139 contigs of which 63 showed homology to the transcripts in the NCBI database. A total of 35 showed homology to plants and 28 had homology to other organisms. A further 56 contigs were ascribed only as hypothetical proteins and four contigs as unknown proteins while 16 contigs showed no homology to any sequence on the NCBI non-redundant database (Table 5). The contigs were grouped into various functional classes and interesting stress response and cell wall associated transcripts are discussed below.

Infected library 2 contained 188 contigs of which 106 showed homology to the transcripts in the NCBI database. A total of 56 showed homology to plants and 50 had homology to other organisms. A further 53 contigs were ascribed only as hypothetical proteins and five contigs as unknown proteins while 24 contigs showed no homology to any sequence on the NCBI non-redundant database (Table 6). The contigs were grouped into various functional classes and interesting stress response and cell wall associated transcripts are discussed below.

© University of Pretoria

2.3.4.2 Transcripts involved in primary metabolism

Various transcripts linked to primary metabolism were discovered in the infected libraries. An interesting fact is that studies on plant–pathogen interactions often utilize pathogen concentrations that are much more than expected to occur in nature to ensure infection takes place. This overloading of the infection mechanism could possibly relate to the up-regulation of primary metabolism genes (Bolton, 2009).

ESTs identified from uninfected avocado roots (0 hr)

There was a significant lack of defence-related transcripts in this library except for the presence of metallothionein. The reason that this defence-related transcript is present in the uninfected library may be due to its role in both metal chaperoning and scavenging of reactive oxygen species in plants (Mir *et al*., 2004). The tolerant avocado rootstock may also have a high basal transcription level of this gene. Indeed we did identify the metallothionein transcript in both infected libraries 1 and 2 indicating that the genes constitutive expression may contribute to its tolerance to *P. cinnamomi*.

Ribosomal RNA transcripts were also found in the sequence data from all three libraries but at a comparatively higher amount in the uninfected library. Fifteen percent of the uninfected library contained ribosomal sequence reads whereas only 0.3 % of either infected library contained ribosomal sequence reads. Ribosomal RNA was not desired in the final sequencing reaction but due to the abundance of ribosomal RNA present in the sample some rRNA was sequenced even though the sample was subjected to mRNA purification.

© University of Pretoria

ESTs identified from avocado infected library 1 containing transcripts from 6 & 12 hr after *Pc* infection

The leucine-rich repeat (LRR) resistance proteins (contig 20) contain a nucleotide binding site (NBS) and a LRR domain. They are important in the transduction of plant internal and external signals and are involved in pathogen recognition. The majority of resistance genes encode proteins that contain the NBS and LRR domains (Dodds *et al.*, 2001; Tichtinsky *et al.*, 2003). The *R1* gene identified in potato belongs to a leucine zipper/NBS/LRR class of genes and was linked to the hyper-sensitive resistance response upon infection with *P. infestans.* The gene was identified on potato chromosome V and characterized in the hope of understanding how the resistance protein recognizes the avirulence effectors. Contig20 is best represented by a hit to *Gossypium hirstum* and most probably represents a pattern recognition receptor however it also shows homology to the Cf-5 disease resistance protein from *Arabidopsis thaliana* and we therefore hypothesize that it may indeed be a defence-related protein.

Pantothenate kinase (contig60) is the first enzyme in the biosynthetic pathway of coenzyme-A which causes phophorylation of pantothenate (Vitamin B5) to form 4'-phosphopantothenate (Kupke *et al.*, 2003). The up-regulation of this gene in a defence response was first identified in a study on *Medicago truncatula* roots in response to *Orobanche crenata*, a parasitic plant (Dita *et al.*, 2009). Not much more is known about this transcript in relation to plant defence but it has been linked to lipid synthesis (Tilton *et al.*, 2006).

Both the plastidic adenylate transporter (contig91) and the patatin-like protein 6 (PLA IIB/PLP) (contig96) have been shown to be linked to plant-pathogen interactions but are in need of further investigation. The plastidic adenylate transporter plays a role in transcription and contains a conserved region that groups it with ATP/ADP transporters. This adenylate

© University of Pretoria

transporter has been implicated in starch biosynthesis in potatoes (Tjaden *et al*., 1998) and the anti-sensing of the transporter gene in potato tubers conferred increased resistance to pathogens such as *P. infestans* (Conrath *et al*., 2003). The identification of this transcript in the avocado infection library 1 implies that this transcript is still being actively transcribed and perhaps down-regulated at a later time point. Anti-sensing of the gene may increase resistance in the already tolerant R0.09 avocado rootstock. The patatin-like protein 6 (PLA IIB/PLP6) has a nutrient reservoir activity and is involved in the lipid metabolic process. It has been shown to be induced in the soybean–*Phytophthora sojae* interaction although there is no information as to the role it plays in the defence response (Narayanan *et al*., 2009). Plant defence responses are complex and involve a combinatorial effect of multiple factors involving gene regulation of different cellular processes. Therefore, the up-regulation of these genes can be interpreted as an increase in cellular activity of the avocado roots due to infection by the pathogen.

ESTs identified from avocado infected library 2 containing transcripts from 24, 48 & 72 hr after *Pc* infection

It has been demonstrated that glyceraldehyde-3-phosphate dehydrogenase (G3PDH) plays a role outside of glycolysis and mediates (reactive oxygen species) ROS signalling (Hancock *et al*., 2005). Contig100 has been identified as a G3PDH that specifically plays a role in the response to oxidative stress and hydrogen peroxide. Oxidative stress is induced by a wide range of environmental factors including pathogen invasion resulting in the generation of the hypersensitive reaction. It has been previously observed that under conditions of stress, the amount of G3PDH increased significantly (Bustos *et al*., 2008)

© University of Pretoria

Cytochrome P450 is a very large and diverse super-family of heme-thiolate proteins (Werck-Reichhart & Feyereisen, 2000). One transcript representing a cytochrome P450-like TBP (TATA box binding protein) (contig187) was identified and overrepresented in the infected library 2. The P450 cytochrome is involved in a wide range of biosynthetic reactions, oxidative metabolism and the production of ROS as well as being greatly involved in plant stress responses that include wounding and pathogen infection (Bolwell *et al.*, 1994). The reason for this overrepresentation of this transcript in the avocado root is probably due to the universal nature of the protein in cell metabolism and growth. However, other studies on different host-pathogen interactions have indicated an increase in expression of the transcript in response to biotic pressures on the plants (Coram *et al.*, 2008; Fan *et al.*, 2011; Perrigault *et al.*, 2009; Thomas *et al.*, 2002).

Arguably the most interesting transcripts identified in all three libraries (contig33 of uninfected, contig76 of infected library 1 & contig54 & contig64 of infected library 2) were that of the metallothionein-like proteins. The three transcripts showed homology to metallothionein-like proteins from different species on the NCBI database, each supported by strong E-values. This metallothionein-like protein type two (contig76) was similar to a sequence from *P. americana.* Metallothionein is a plant stress gene responsible for the detoxification (regulation) of heavy metals in cells (John *et al.*, 2009) and plays a role in inorganic ion transport and metabolism. This gene has shown increased expression during a number of abiotic and biotic stresses in *Brassica napus* and *A. thaliana* and has been implicated in detoxification of heavy metals in cells (Butt *et al.*, 1998; John *et al.*, 2009). Metallothioneins have also been shown to inhibit programmed cell death (PCD) and Fumonisin B1-induced root death in *Solanum lycopersicum* infected with *Agrobacterium rhizogenes* (Harvey *et al.*, 2008). Expression of plant metallothionein genes has been reported in leaves and stems, ripening fruits, and wounded

© University of Pretoria

tissues, and has been proposed to function in both metal chaperoning and scavenging of reactive oxygen species (Mir *et al.*, 2004). In the study by Harvey *et al.* (2008) it was also demonstrated that ROS accumulation was significantly reduced under metallothionein over-expression, validating its function in ROS scavenging. Our investigations of metallothionein expression in avocado roots have revealed that the number of metallothionein reads was the lowest in the uninfected library and highest in the infected library 1 (6 & 12 hr). Metallothionein down-regulation in the R0.09 avocado rootstock is expected since the rootstock exhibits tolerance to the pathogen.

2.3.4.3 Transcripts involved in stress-related responses

Stress-related proteins are expressed in a plant system when the plant encounters biotic or abiotic external stresses. These stress responses allow the plant cell to protect itself from cell damage by unfavourable environment conditions and pathogens. Stress leads to an increase in metabolic function due to the need for transcription and translation of defence-related genes.

ESTs identified from avocado infected library 1 containing transcripts from 6 & 12 hr after *Pc* infection

Some of the defence-related transcripts found in the avocado infected library 1 were the alpha-amylase inhibitor-like precursor (contig34), bZIP transcription factor bZIP80 (contig47), a homolog to the pathogenesis-related protein *PsemI* (contig106), dehydrin (contig109) and a 70 kDa peptidyl-prolyl isomerase (contig111).

© University of Pretoria

The α-amylase inhibitor-like precursor is a protein that inhibits α-amylase and plays a role in the stress response of plants against insects. An important feature of α-amylase inhibitors is that their inhibitory activity can be species-specific, such as members of the cereal family having amylase inhibitors that are active against insect a-amylases (Chagolla-Lopez *et al*., 1994). It is not clear what effect this protein may have in the avocado-*P. cinnamomi* interaction.

Proteins with bZIP domains are present in all eukaryotes. Members of the bZIP family play various roles which include pathogen defence (Jakoby *et al*., 2002). The bZIP80 (contig47) transcription factor that showed homology to the sequence generated from avocado has been demonstrated to play a role in abscisic acid (ABA) signalling and conferred salt and freezing tolerance (Liao *et al*., 2008). The identification of this gene implies the activation of the ABA signalling pathway in the avocado root in response to infection by *P. cinnamomi* which is supported by the fact that a bZIP transcription factor was also found to be induced in the soybean– *P. sojae* interaction (Narayanan *et al*., 2009).

Contig106 was identified as a homolog to the pathogenesis-related protein *PsemI* and hypothesized to be a PR10 protein. PR-10 proteins have been shown to be induced by elicitors, pathogens and metal stress and may have ribonuclease-like activity. *PsemI* is a PR10 protein and has been shown to be up-regulated in Douglas-fir and has been considered as being used as a marker in screening for *Phellinus weirii* resistance in Douglas-fir (Ekramoddoullah *et al*., 2000). Another *PR-10* gene (*CaPR-10*) from *Capsicum annuum* demonstrated inhibition of *Phytophthora capsici* growth (Park *et al*., 2004). Contig109 also showed a high degree of homology to PR-10 proteins. It was identified as being a dehydrin, which is a drought-induced protein, and had a 70% sequence homology to *AhPR10* from *Arachis hypogaea*. AhPR10's antifungal activity was proven against *Fusarium oxysporum* and *Rhizoctonia solani* (Chadha & Das, 2006; Pnueli *et al*., 2002). The results imply the role of contig106 and contig109 in defence

© University of Pretoria

against *P. cinnamomi* and the potential use of contig106 in screening for PRR tolerance in avocado.

ESTs identified from avocado infected library 2 containing transcripts from 24, 48 & 72 hr after *Pc* infection

Contig7, Contig163 and Contig176 have shown much similarity to the translationally controlled tumour protein (TCTP)-like protein. It is highly conserved and expressed in eukaryotes and has been implicated in important cellular processes, such as cell growth, cell cycle progression, and in the protection of cells against various stress conditions and apoptosis (Bommer & Thiele, 2004). It can be speculated that tumour proteins may be related to stress response in plants where a study demonstrated differential display of TCTP proteins in response to *Ralstonia solanacearum* infection in *Nicotiana tabacum* (Kiba *et al.*, 2007).

Ascorbate peroxidase (APX) (contig151) takes part in many metabolic processes, including chloroplastic, mitochondrial, and plasma membrane-linked electron transport systems. APX scavenges reactive oxygen species such as hydrogen peroxide in a bid to protect against the cellular damage (Asada, 1992; Noctor & Foyer, 1998). Hydrogen peroxide plays an important role in the activation and propagation of pathogen-induced programmed cell death (PCD) in plants and should not be inactivated by APX in the avocado root under *P. cinnamomi* infection. Research has shown that the expression of cytosolic APX is post-transcriptionally suppressed during pathogen infection (Mittler *et al.*, 1998). This information implies that increase in transcripts of APX in the infected avocado library may differ considerably from the amount of active protein. Decreased APX activity would mean an increase in ROS and subsequent PCD.

© University of Pretoria

PR5 proteins are described as being inducible by biotic stresses and further linked to increased pathogen resistance (Filippov *et al.*, 2005). Two *PR5* transcripts were found in the infected avocado library 2. Thaumatin (contig11) associated with the salicylic acid (SA) pathway (Liu *et al.*, 2010; Wang *et al.*, 2010), and the thaumatin-like protein (contig83) which is usually induced by viral, bacterial and fungal infection (Liu *et al.*, 2010). Thaumatin expression may contribute to the overall tolerance of the plant in response to the pathogen which was demonstrated when transgenic potato expressing the thaumatin II gene , showed increased partial resistance to *P. infestans* (Filippov *et al.*, 2005). The *PR5* genes' activity in the SA-dependant pathway has been shown to be effective against biotrophs (Thomma *et al.*, 1998). The thaumatin-like protein was found to have antimicrobial properties alone or in a combination (Uknes *et al.*, 1992) and although thaumatin-like proteins show strong antifungal activity, their antifungal mode of action is still not completely understood (Monteiro *et al.*, 2003).

The universal stress protein (USP) (contig175) was originally described in prokaryotic organisms. It has been documented in rice as playing a role in ethylene mediated stress adaptation (Sauter *et al.*, 2002). In a study on *Eucalyptus grandis* infected with *Puccinia psidii* Winter, it was found that USP was more expressed in susceptible material and it is hypothesized that it forms part of a general stress response (Moon *et al.*, 2007). This protein may also play a general defence role in the avocado-*P. cinnamomi* interaction.

Contig73 is homologous to the oxysterol-binding protein (OBP). The up-regulation of the OBP in avocado is possibly related to the *P. cinnamomi* infection and its associated elicitins. OBPs are nuclear receptors that also function as transcription factors. In the study of potato-*P. infestans* interaction the OBP was up-regulated and is thought to be part of a non-specific defence pathway (Avrova *et al.*, 2004). The exact role of OBPs is unknown but there is a connection between the sterol binding capacity of *Phytophthora*-secreted elicitins and OBPs and their ability

© University of Pretoria

to activate defence responses. *Phytophthora* species produce elicitins which are small cysteine-rich lipid-binding proteins secreted by oomycetes. These elicitins bind to sterols produced by the plant. The sterol-elicitin complex then triggers a cellular response (Blein *et al.*, 2002; Osman *et al.*, 2001).

A putative cysteine proteinase precursor (contig53) and cysteine proteinase (contig169) have been identified in the avocado infected library 2. Cysteine proteinases play essential roles in developmental and physiological processes in plants (Ho *et al.*, 2000). The cysteine proteinase precursor contains a cathepsin propeptide inhibitor domain and possibly regulates programmed cell death. The cysteine proteinase identified in the avocado roots could be linked to the hypersensitive response generated by *P. cinnamomi* since cysteine proteinases are mainly involved in the reactions leading to programmed cell death (Ho *et al.*, 2000).

2.3.4.4 Transcripts involved in the plant cell wall

The synthesis of cell wall proteins is stimulated during development but also under stress conditions. This is one of the first lines of defence against pathogen infection but is also activated after infection has occurred. The infiltration of *P. cinnamomi* may activate cell wall strengthening genes in avocado roots as a mechanical defence against the pathogen or to repair damage caused by infection.

© University of Pretoria

ESTs identified from avocado infected library 1 containing transcripts from 6 & 12 hr after *Pc* infection

The putative identification of contig88 is a seven transmembrane protein- *Mlo6.* MLO proteins consist of seven membrane spanning domains and function as G protein coupled receptors. They relate extracellular signals into-intra cellular responses and modulate defence responses. In barley the *Mlo* gene modulates defence responses to *Blumeria graminis* (Büschges *et al*., 1997; Devoto *et al*., 2002). Barley infected with the biotrophic powdery mildew fungus *Blumeria graminis* f. sp. *hordei* has shown that actin cytoskeleton polarization is modulated by the receptor-like transmembrane protein MLO at sites of attempted penetration (Büschges *et al*., 1997) with the frequency of successful defence correlating with strong actin focus towards sites of attempted penetration and *mlo* activity (Opalski *et al*., 2005). However in a study of barleys response to a necrotrophic pathogen- *Bipolaris sorokiniana* it was shown that a *mlo* compromised pathway resulted in the plant exhibiting higher susceptibility while the *mlo* wild type expression results in resistance (Kumar *et al*., 2001). Again when barleys response to the hemibiotroph *Magnaporthe grisea* (which is the causal agent of rice blast) was investigated, it was observed that increased susceptibility is achieved with mutation of *mlo* (Jarosch *et al*., 1999). These studies suggest that *mlo* may modulate different resistance mechanisms against biotrophic and necrotrophic fungi, which is supported by an independent hypothesis that wild-type *mlo* may modulate different resistance mechanisms effective against biotrophic and necrotrophic fungi, in opposite directions (Kumar *et al*., 2001; Piffanelli *et al*., 2002). It is a possibility that *mlo* mutants assist necrotrophic pathogens in their lifestyles by triggering mesophyll cell death upon inoculation. *Mlo* expression is a prerequisite for resistance to necrotrophic penetration, but its expression increases susceptibility to biotrophic organisms. The role of *mlo* in the avocado-*P. cinnamomi* interaction is unknown but *mlo* may also function as a

© University of Pretoria

suppressor of cell death and there is evidence that it plays a role in mesophyll cell survival (Devoto *et al.*, 2002; Kumar *et al.*, 2001).

ESTs identified from avocado infected library 2 containing transcripts from 24, 48 & 72 hr after *Pc* infection

Contig108 had the highest homology to the profilin-like protein. Profilins are a ubiquitous family of proteins that control actin polymerization in eukaryotic cells and can either promote or prevent actin polymerization (Sun *et al.*, 1995). The actin cytoskeleton plays a critical role in cell division and expansion. After fungal infection thickening of the cell wall has been observed, this involves the reorientation of actin filaments as a defence response in order to prevent pathogen ingress (Schmelzer, 2002). In the parsley-*Phytophthora infestans* interaction it was demonstrated that profilin localized to the site beneath cell wall penetration subtending the oomycetous appresorium at the plasma membrane (Schütz *et al.*, 2006). This profilin accumulation at sites of highly dynamic actin filaments draws the conclusion that profilin accumulation is linked with increased cell division and expansion in response to fungal penetration. The presence of the profilin-like transcript in the infection library implicates actin polymerization and could indicate that the avocado root activates cell wall thickening in response to pathogen penetration.

A PDR-type ABC (ATP-binding cassette) transporter 2 (contig167) was identified in infected library 2. The pleiotropic drug resistance-like (PDR) family of ABC transporters participates directly in the transport of molecules across membranes and is one of the three best characterized ABC subfamilies which include multidrug resistance (MDR), and multidrug resistance-associated protein (MRP) subfamilies (Jasinski *et al.*, 2003). The PDR family of transporters is distinguished by their association with the transport of antifungal agents and may

© University of Pretoria

play a role in plant-fungal interactions (van den Brûle & Smart, 2002). PDR-type ABC transporters have been shown to mediate the cellular uptake of abscisic acid (ABA) and stimulate the response of ABA receptive genes (Kang *et al.*, 2010). Although PDR-type ABC transporters regulate the uptake of ABA their expression may not always be induced by ABA. It was demonstrated that the *NtPDR1* gene in tobacco was induced by methyl jasmonic acid (MeJA) (Sasabe *et al.*, 2002). Sasabe *et al.* (2002) also demonstrated that microbial elicitors induce the expression of a PDR-type ABC transporter. Tolerant avocado rootstocks such as R0.09 may rely on this PDR-ABC transporter in the defence induction in response to *P. cinnamomi*.

Three of the five most abundant cell wall proteins are glycine-rich proteins (GRPs), extensins and proline-rich proteins (Showalter *et al.*, 1996). We have identified a glycine-rich protein (contig64) and putative extensins (conti137 & contig139 of infected library 1 & contig94 of infected library 2) that were induced in the avocado-*P. cinnamomi* interaction. These types of proteins have been shown to be components of the cell walls of many plants and their synthesis may also be linked to part of the plant's defence mechanism (Mousavi & Hotta, 2005). GRPs are mostly regulated in the developmental stages of the plant but also during pathogen infection with studies showing that *GRP* genes are upregulated in response to external biotic and abiotic stress (Mousavi & Hotta, 2005). The GRP from avocado (contig64) contains a nucleic binding region and plays a role in signal transduction according to the GO and KOG databases. Extensins are a family of flexuous, rod-like, hydroxyproline-rich glycoprotein's (HRGPs) of the plant cell wall. When the activity of extensin was studied in response to fungal activity there was an increase in extension in melon plants infected with the anthracnose fungus (Esquerre-Tugaye & Mazau, 1974). Another cell wall protein known as ENOD2 (Contig25) was also identified. The role of this protein is still debated amongst the scientific community because of

© University of Pretoria

the genes' transcription in non-nodulating woody plants (Foster *et al.*, 2000) but according to the gene ontology classification, the domains identified indicate that the protein contributes to the structural integrity of the cell wall. It was demonstrated in soybean that levels of ENOD2 increased due to drought-stress, suggesting that this gene is involved in stress responses (Skøt *et al.*, 1996). The presence of these transcripts can be interpreted as a form of mechanical defence by the avocado root in response to *P. cinnamomi.*

Lignin biosynthesis is essential for the structural integrity of plant cell walls but has also been described as a means of defence against pathogen infiltration (Vance *et al.*, 1980). We have identified two transcripts from infected library 1 (contig38 & contig70) and one transcript from infected library 2 (contig43) related to lignin production. Caffeoyl-CoA 3-O-methyltransferase (contig38) belongs to the transferase family and functions in O-methyltransferase activity. The down-regulation of this enzyme results in decreased lignin content (Guo *et al.*, 2001). The 4-coumarate-CoA ligase-like protein (4CL) (contig70) is a key enzyme of general phenylpropanoid metabolism. The protein provides precursors for lignin biosynthesis (Stuible *et al.*, 2000). Cinnamate 4-hydroxylase (contig43) catalyzes the conversion of cinnamate into 4-hydroxy-cinnamate, a key reaction of the phenylpropanoid pathway. It is responsible for lignin biosynthesis and plant defence responses (Bell-Lelong *et al.*, 1997; Schoch *et al.*, 2002). Previous studies showed that defence lignin was found to be significantly different from lignin in vascular tissues, this suggests that lignin biosynthesis is differentially regulated (Bhuiyan *et al.*, 2009).

© University of Pretoria

2.3.5 Sequence alignments of selected defence-related ESTs

Defence-related genes were identified among various functional classes in both infected libraries 1 and 2 (Table 3). Transcripts such as thaumatin, a thaumatin-like protein, a universal stress gene, profilin-like gene, a MLO transmembrane encoding gene, the pathogenesis-related protein PR10, the metallothionein-like type 2 EST, the LRR resistance gene and the cytochrome P450-like transcript were chosen for expression analysis. Before expression analysis could be done the sequences were searched for conserved domains. The alignments of nine defence genes were analysed to assess the conservation of amino acids as well as to investigate the portion of genes predominantly sequenced.

The thaumatin EST (contig11 of infected library 2) was translated in the -2 reading frame and consisted of 80 amino acid residues. It aligned to the 3' end of the amino acid sequence of other thaumatin genes from *Vitis vinifera, V. riparia, V. riparia* crossed with *V. berlandieri* and *Populus tomentosa.* The EST showed an 87% homology to a 247 amino acid *V. riparia thaumatin* gene (GI:5881239) with a few changes toward the end of the amino acid sequence (Figure 5). The translated amino acid sequence was associated with the thaumatin family from the SMART database (SMART accession SM00205) and grouped to the super-family GH64-TLP-SF (glycoside hydrolase family 64 (beta-1,3-glucanases which produce specific pentasaccharide oligomers) and thaumatin-like proteins) on the NCBI conserved domains database.

The thaumatin-like protein (contig83 of infected library 2) consisted of 74 amino acid residues and was translated in the +2 reading frame. This EST showed a 70% similarity with an *A. thaliana* thaumatin-like protein (GI:15241688) with an 87% query coverage. Other sequences that the EST showed close homology to included *V. vinifera, Oryza sativa* and *Nicotiana tabacum.* The amino acid sequence shared the same super-family and family as the thaumatin

© University of Pretoria

transcript as well as showing similarity to the TLP-PA (allergenic/antifungal thaumatin-like proteins: plant and animal homologs- cd09218) family (Figure 6).

Contig 175 of infected library 2 was identified as a universal stress protein and consisted of 91 amino acid residues that were translated in the +3 reading frame. It showed a 89% identity over 80% of the query sequence with *Cicer arietinum* (GI:45720184). The amino acid sequence of this EST contains eight conserved residues that comprise the ligand binding site of the universal stress protein family (cd00293) from residue 25 to 55 (Figure 7). The fragment falls within the AANH (Adenine nucleotide alpha hydrolases super-family- cd0292).

The metallothionein-like type 2 EST (contig64 of infected library 1) consisted of 54 amino acid residues translated in the -1 reading frame. This transcript showed 98% homology to a *P. americana* fragment in the NCBI database (GI:7242691). The amino acid sequence grouped into the metallothionein protein family (pfam01439) under the metallothio_2 super-family (cl03212) (Figure 8).

The profilin-like EST (contig108 of infected library 2) consisted of 42 amino acid residues translated in the +2 reading frame. It showed homology to allergen genes from *Hevea brasiliensis* and *Mercurialis annua* but had a 100% homology to the profilin-like protein from *Cinnamomum camphora* (GI:84682949). The amino acid sequence also contained six of 11 conserved residues of the actin interaction site as well as two of the six residues of the poly-proline binding site (Figure 9). This EST belonged to the profilin family (pfam00235) under the PROF super-family (cl00123).

The MLO transmembrane transcript was translated in the +1 reading frame and consisted of 83 amino acid residues which showed a 70% homology to the barley *mlo* defence gene (Figure 10). Other protein sequences that showed close homology to the amino acid sequence of contig

© University of Pretoria

88 from infected library 1 were from *V. vinifera, Ricinus communis,* and *Malus toringoides*. The amino acid query was grouped into the Mlo protein family (pfam03094) in the Mlo super-family (cl03887).

The pathogenesis-related protein PR10 EST (contig106 of infected library 1) showed 46% homology to the *PsemI* pathogenesis-related protein from *Pseudotsuga menziesii* (GI:11080640) with a 96% coverage of the query sequence. Although the homology is low the total coverage is very high and we therefore hypothesize that this EST belongs to the PR10 family. The EST consisted of 79 amino acid residues that were translated in the +2 reading frame. Other species that shared close homology to the query sequence were PR10 transcripts of *Pinus monticola* and *Elaeis guineensis*. Twenty two of the 37 conserved amino acid residues that comprise the characteristic hydrophobic ligand binding region of this family are present in the EST (Figure 11). The Bet_v1-like domain (cd07816) that the sequence groups under contains PR10 proteins under the SRPBCC (START/ RHO_alpha_C /PITP /Bet_v1/ CoxG/ CalC) ligand-binding super-family (cl14643).

The LRR resistance transcript consists of 80 amino acid residues when translated in the +1 reading frames. It showed an 80% homology to leucine-rich repeat genes from *A. thaliana, Gossypium hirsutum, Glycine max* and a *Brassica rapa* subspecies, but did not contain any functional domains (Figure 12). The transcript also showed homology to the Cf-5 disease resistance protein from *Arabidopsis thaliana.* Leucine-rich repeats are important in the transduction of plant internal and external signals (Tichtinsky *et al.*, 2003).

The cytochrome P450-like protein consisted of 263 amino acid residues and was translated in the -1 reading frame. It shares a 75% homology with *Zea mays* (GI:7489812) cytochrome P450 mono-oxygenase as well as a 76% homology with cytochrome P450 like TBP from *N. tabacum*.

© University of Pretoria

The amino acid sequence contains four of the 19 conserved residues of the NAD(P) binding site (Figure 13).

2.3.6 Species similarity between avocado and other plants

We observed significant sequence homology between *V. vinifera* (grape) and avocado when the species origin of the sequence similarity was investigated. The top three represented species according to amino acid homology on the NCBI were *V. vinifera, A. thaliana* and *O. sativa,* with *V. vinifera* having the majority of the hits in all three libraries. Twenty two percent of sequences showed homology to *V. vinifera* sequences with 7.5 % belonging to *A. thaliana* and 7.8 % of sequences to *O. sativa*. Homology to *P. americana* was found in only 1 % of sequences (4/371) (Table 8). Only two genes were represented by the 1 % in which the metallothionein transcript featured three times and fructose- bisphosphate aldolase featured once. Grape vine featured among the top ten homologous hits of every contig that was annotated. Thirty seven percent of the annotated contigs were represented by various plant species such as *Prunus armeniaca*, *Solanum tuberosum*, *Hevea brasiliensis* with the variety of plant species not biased to any particular family or order. The majority of the species similarities relate to a large variety of plants that have been collectively categorised as other.

© University of Pretoria

2.4 Conclusion

This study generated sequence data for the roots of a highly tolerant avocado rootstock in response to *P. cinnamomi* and identified defence-related genes. This was successfully achieved by generating the first 2.0 Mb of transcriptomic data and identifying 367 novel transcripts for avocado. The importance of avocado as fruit crop has already been demonstrated (Chanderbali *et al.*, 2008) and the need to develop molecular tools to improve the crop have been expressed and initiated (Lavi *et al.*, 1994; Lavi *et al.*, 1991; Mhameed *et al.*, 1995; Mhameed *et al.*, 1997; Sharon *et al.*, 1997). The method chosen to sequence the ESTs from the avocado roots was 454 pyrosequencing due to the limitations of other sequencing platforms. High throughput sequencing has been used for the *de novo* transcriptome assembly of non-model organisms with the pyrosequencing technology being the technology of choice due to its longer reads, ease of assembly, and ease of library preparation and in terms of logistics the most economical and effective (Collins *et al.*, 2008; Meyer *et al.*, 2009). The high data output and rapid detection of high-throughput sequencing does prove to be better than EST sequencing (Weber *et al.*, 2007).

In total the 2 Mb that was generated from the sequencing run was close to its theoretical maximum of 2.5 Mb since a single lane of the GS FLX platform was used. The 0.5 Mb loss is due to failure of wells and removal of tag sequences of the three libraries. Taking both these allowances into consideration we generated a large amount of valuable sequence information. The uninfected library generated 0.3 Mb, the infected library 1, 0.7 Mb and infected library 2 generated 1.0 Mb. The 9953 reads had an average read length of just under the threshold of 250 bp. The uneven distribution of the reads over the three response libraries is in no way an identification of the transcript abundance but rather due to random sampling of the libraries prior to sequencing.

© University of Pretoria

It should be noted that the data represented in the GO, KOG graphs consists of combining the contigs of the three cDNA libraries. Gene ontology indicates that the vast majority of transcripts are unknown with the top ten mostly populated functional groupings according to the GO database beginning with the unknown functional class followed by other, cellular component, biological process, stress response, ribosome structure, cell wall related, protein binding, mitochondrion and ATP binding. According to the KOG database the majority of sequence data matched categories of general function prediction only, meaning that transcripts were not unequivocally assigned to a certain group. The KOG database revealed that the top three classes that the contigs grouped into were the general function prediction class followed by signal transduction and unknown function. The information described in this study indicates the distribution of transcripts for our sequencing efforts and the data is not a complete description of the transcriptome. As can be seen in other transcriptome studies the large amount of transcripts that group under unknown function is also expected (Coram *et al.*, 2008). We expected a large amount of data to remain un-annotated due to the limited nature of our sequencing efforts, the absence of a closely related genome sequence and due to the lack of sequence available for *P. americana* in the public domain.

The 371 contigs generated contained genes from a vast array of processes with only two genes being previously identified in avocado (Fructose-bisphosphate aldolase and metallothionein type-II protein) with the remaining 367 having not been previously identified indicating the usefulness of the method to identify genes in a non-model organism (Priest *et al.*, 2010). Interestingly, various transcripts identified in this sequencing effort have been previously linked to plant-*Phytophthora* species interactions in other studies. Contig 91 (plastidic adenylate transporter) and 96 (patatin-like protein 6 (PLA IIB/PLP)) of infected library 1 have been linked to host responses to *P. infestans* and *P. sojae* infection (Conrath *et al.*, 2003; Narayanan *et al.*,

© University of Pretoria

2009) while contig 11 (thaumatin), 73 (oxysterol-binding protein) and 108 (profilin-like protein) of infected library 2 have been linked to *P. infestans* interactions (Avrova *et al*., 2004; Filippov *et al*., 2005; Schütz *et al*., 2006). Numerous lignin associated transcripts were also identified (contig 38 and 70 from infected library 1 and contig 43 of infected library 2) which support the data of other studies indicating the up-regulation of the lignin synthesis pathways (Rookes *et al*., 2008).

The alignments of selected defence-related genes show homology to the 3' end of the amino acid sequence which has been documented before (Bainbridge *et al*., 2006). This indicates that the sequencing run had a bias towards the 3' end of the transcripts. This can be explained by the method of purification or mRNA from total RNA because the Oligotex beads selectively bind to the poly(A) tail of transcripts along with the Roche cDNA synthesis system employing oligo-dT primers (Durban *et al*., 2011). A 3'-enrichment of transcripts can be avoided by either replacing oligo-dT primers with random primers or by the normalization of cDNA libraries before sequencing (Vega-Arreguin *et al*., 2009; Wall *et al*., 2009). Other studies however, indicate that strategies such as random priming are not needed to generate a good overall representation of transcripts (Weber *et al*., 2007). This leads us to believe that there are many factors (such as quality of RNA, mRNA selection criteria, cDNA synthesis as well as the pyrosequencing reaction itself) that influence the length of sequence reads generated.

A large proportion of the avocado transcripts showed significant amino acid sequence homology to *V. vinifera* on the NCBI. Species such as *A. thaliana,* and *O. sativa* were also screened for homology to *P. americana* but there were only a small number of transcripts that associated with these organisms. The low number of hits on any avocado sequence data (only 1%) was expected, highlighting the lack of genetic data present for the crop and also because the current

© University of Pretoria

sequence data available is related to fruiting and flowering characteristics and will not necessarily be found in the root material.

This is the first study to sequence the transcriptome of the avocado roots infected with *P. cinnamomi.* The sequence data generated has thus far lead to the identification of putative defence related genes. These transcripts have not yet been quantitatively linked to *P. cinnamomi* infection and will be analysed using qRT-PCR at different time intervals. The sequences generated will help in the elucidation of the avocado transcriptome and also form a basis for other studies to be conducted with the development of microarray analysis for avocado root expression becoming a reality.

2.5 References

Altschul S. F., Gish W., Miller W., Myers E. W., Lipman D. J. (1990) Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Asada K. (1992) Ascorbate peroxidase – a hydrogen peroxide-scavenging enzyme in plants. *Physiologia Plantarum* **85**, 235-241.

Ashworth V. E. T. M., Kobayashi M. C., De La Cruz M., Clegg M. T. (2004) Microsatellite markers in avocado (*Persea americana* Mill.): development of dinucleotide and trinucleotide markers. *Scientia Horticulturae* **101**, 255-267.

Avrova A. O., Taleb N., Rokka V. M., Heilbronn J., Campbell E., Hein I., Gilroy E. M., Cardle L., Bradshaw J. E., Stewart H. E., Fakim Y. J., Loake G., Birch P. R. J. (2004) Potato oxysterol binding protein and cathepsin B are rapidly up-regulated in independent defence pathways that distinguish *R* gene-mediated and field resistances to *Phytophthora infestans*. *Molecular Plant Pathology* **5**, 45-56.

© University of Pretoria

Bainbridge M., Warren R., Hirst M., Romanuik T., Zeng T., Go A., Delaney A., Griffith M., Hickenbotham M., Magrini V., Mardis E., Sadar M., Siddiqui A., Marra M., Jones S. (2006) Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *BMC Genomics* **7**, 246.

Bell-Lelong D. A., Cusumano J. C., Meyer K., Chapple C. (1997) Cinnamate-4-hydroxylase expression in *Arabidopsis*: regulation in response to development and the environment. *Plant Physiology* **113**, 729-738.

Ben-Ya'acov A., Michelson E. (1995) Avocado rootstocks. *Horticultural Reviews* **17**, 381-429.

Bhuiyan N. H., Selvaraj G., Wei Y., King J. (2009) Role of lignification in plant defence. *Plant Signaling & Behavior* **4**, 158-159.

Blein J.-P., Coutos-Thévenot P., Marion D., Ponchet M. (2002) From elicitins to lipid-transfer proteins: a new insight in cell signalling involved in plant defence mechanisms. *Trends in Plant Science* **7**, 293-296.

Bolton M. D. (2009) Primary metabolism and plant defence—Fuel for the fire. *Molecular Plant-Microbe Interactions* **22**, 487-497.

Bolwell G. P., Bozak K., Zimmerlin A. (1994) Plant cytochrome P450. *Phytochemistry* **37**, 1491-1506.

Bommer U.-A., Thiele B.-J. (2004) The translationally controlled tumour protein (TCTP). *The International Journal of Biochemistry & Cell Biology* **36**, 379-385.

Borrone J. W., Schnell R. J., Violi H. A., Ploetz R. C. (2007) Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. *Molecular Ecology Notes* **7**, 439-444.

© University of Pretoria

Büschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., van Daelen R., van der Lee T., Diergaarde P., Groenendijk J., Töpsch S., Vos P., Salamini F., Schulze-Lefert P. (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* **88**, 695-705.

Bustos D. M., Bustamante C. A., Iglesias A. A. (2008) Involvement of non-phosphorylating glyceraldehyde-3-phosphate dehydrogenase in response to oxidative stress. *Journal of Plant Physiology* **165**, 456-461.

Butt A., Mousley C., Morris K., Beynon J., Can C., Holub E., Greenberg J. T., Buchanan-Wollaston V. (1998) Differential expression of a senescence-enhanced metallothionein gene in *Arabidopsis* in response to isolates of *Peronospora parasitica* and *Pseudomonas syringae*. *Plant Journal* **16**, 209-221.

Chadha P., Das R. (2006) A pathogenesis related protein, AhPR10 from peanut: an insight of its mode of antifungal activity. *Planta* **225**, 213-222.

Chagolla-Lopez A., Blanco-Labra A., Patthy A., SÃ¡nchez R., Pongor S. (1994) A novel alpha-amylase inhibitor from amaranth (*Amaranthus hypocondriacus)* seeds. *Journal of Biological Chemistry* **269**, 23675-23680.

Chanderbali A. S., Albert V. A., Ashworth V. E. T. M., Clegg M. T., Litz R. E., Soltis D. E., Soltis P. S. (2008) *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. *BioEssays* **30**, 386-396.

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

Chen H., Ashworth V. E. T. M., Xu S., Clegg M. T. (2007) Quantitative genetic analysis of growth rate in avocado. *Journal of the American Society for Horticultural Science* **132**, 691-696.

© University of Pretoria

Clegg M. T., Davis J. W. (1989) Molecular genetics of avocado. *California Avocado Society Yearbook* **73**, 59-61.

Clegg M. T., Henderson D., Durbin M. (1992) The development of molecular markers in avocado. *Proceedings of Second World Avocado Congress 1992, 573.*

Clegg M. T., Kobayashi M., Lin J. Z. (1999) The use of molecular markers in the managment and improvement of avocado (*Persea americana* Mill.). *Revista Chapingo Serie Horticultura* **5**, 227-231.

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

Collins L. J., Biggs P. J., Voelckel C., Joly S. (2008) An approach to transcriptome analysis of non-model organisms using short read sequences. *Genome Informatics* **21**, 3-14.

Conrath U., Linke C., Jeblick W., Geigenberger P., Quick W. P., Neuhaus H. E. (2003) Enhanced resistance to *Phytophthora infestans* and *Alternaria solani* in leaves and tubers, respectively, of potato plants with decreased activity of the plastidic ATP/ADP transporter. *Planta* **217**, 75-83.

Coram T. E., Wang M., Chen X. (2008) Transcriptome analysis of the wheat–*Puccinia striiformis* f. sp. *tritici* interaction. *Molecular Plant Pathology* **9**, 157-169.

Devoto A., Hartmann H. A., Piffanelli P., Elliott C., Simmons C., Taramino G., Goh C.-S., Cohen F. E., Emerson B. C., Schulze-Lefert P., Panstruga R. (2002) Molecular phylogeny and evolution of the plant-specific seven-transmembrane MLO family. *Journal of Molecular Evolution* **56**, 77-88.

Diatchenko L., Lau Y. F., Campbell A. P., Chenchik A., Moqadam F., Huang B., Lukyanov S., Lukyanov K., Gurskaya N., Sverdlov E. D., Siebert P. D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA

© University of Pretoria

probes and libraries. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6025-6030.

Dita M. A., Die J. V., Roman B., Krajinski F., Kuster H., Moreno M. T., Cubero J. I., Rubiales D. (2009) Gene expression profiling of *Medicago truncatula* roots in response to the parasitic plant *Orobanche crenata*. *Weed Research* **49**, 66-80.

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

Dodds P. N., Lawrence G. J., Ellis J. G. (2001) Six amino acid changes confined to the leucine-rich repeat β-strand/ β-turn motif determine the difference between the P and P2 rust resistance specificities in Flax. *Plant Cell* **13**, 163-178.

Douhan G., Fuller E., McKee B., Pond E. (2011) Genetic diversity analysis of avocado (*Persea americana* Miller) rootstocks selected under greenhouse conditions for tolerance to phytophthora root rot caused by *Phytophthora cinnamomi*. *Euphytica* **182**, 209-217.

Durban J., Juarez P., Angulo Y., Lomonte B., Flores-Diaz M., Alape-Giron A., Sasa M., Sanz L., Gutierrez J., Dopazo J., Conesa A., Calvete J. (2011) Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. *BMC Genomics* **12**, 259.

Ekramoddoullah A. K. M., Xueshu Y., Rona S., Arezoo Z., Doug T. (2000) Detection and seasonal expression pattern of a pathogenesis-related protein (PR-10) in Douglas-fir (*Pseudotsuga menziesii*) tissues. *Physiologia Plantarum* **110**, 240-247.

Esquerre-Tugaye M. T., Mazau D. (1974) Effect of a fungal disease on extensin, the plant cell wall glycoprotein. *Journal of Experimental Botany* **25**, 509-513.

© University of Pretoria

Fan F., Li X.-W., Wu Y.-M., Xia Z.-S., Li J.-J., Zhu W., Liu J.-X. (2011) Differential expression of expressed sequence tags in alfalfa roots under aluminium stress. *Acta Physiologiae Plantarum* **33**, 539-546.

Filippov A., Kuznetsova M., Rogozhin A., Spiglazova S., Smetanina T., Belousova M., Kamionskaya A., Skryabin K., Dolgov S. (2005) Increased resistance to late blight in transgenic potato expressing thaumatin II gene. *Ninth workshop of an European Network for development of an integrated control strategy of potato late blight*, 263-267.

Foster C. M., Horner H. T., Graves W. R. (2000) Accumulation of *ENOD2*-like transcripts in non-nodulating woody papilionoid legumes. *Plant Physiology* **124**, 741-750.

Gachon C., Mingam A., Charrier B. (2004) Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany* **55**, 1445-1454.

García-Pineda E., Benezer-Benezer M., Gutiérrez-Segundo A., Rangel-Sánchez G., Arreola-Cortés A., Castro-Mercado E. (2010) Regulation of defence responses in avocado roots infected with *Phytophthora cinnamomi* (Rands). *Plant and Soil* **331**, 45-56.

Giblin F., Pegg K., Willingham S., Anderson J., Coates L., Cooke T., Dean J., Smith L. (2005) *Phytophthora* Revisited. In: *New Zealand and Australia Avocado Grower's Conference*, Tauranga, 1-7.

Guo D., Chen F., Inoue K., Blount J. W., Dixon R. A. (2001) Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa: impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell* **13**, 73-88.

Hancock J. T., Henson D., Nyirenda M., Desikan R., Harrison J., Lewis M., Hughes J., Neill S. J. (2005) Proteomic identification of glyceraldehyde 3-phosphate dehydrogenase as an inhibitory target of hydrogen peroxide in *Arabidopsis*. *Plant Physiology and Biochemistry* **43**, 828-835.

© University of Pretoria

Harvey J., Lincoln J., Gilchrist D. (2008) Programmed cell death suppression in transformed plant tissue by tomato cDNAs identified from an *Agrobacterium rhizogenes*-based functional screen. *Molecular Genetics and Genomics* **279**, 509-521.

Ho S.-L., Tong W.-F., Yu S.-M. (2000) Multiple mode regulation of a cysteine proteinase gene expression in rice. *Plant Physiology* **122**, 57-66.

Jakoby M., Weisshaar B., Dröge-Laser W., Vicente-Carbajosa J., Tiedemann J., Kroj T., Parcy F. (2002) bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science* **7**, 106-111.

Jarosch B., Kogel K.-H., Schaffrath U. (1999) The ambivalence of the barley *Mlo* locus: Mutations conferring resistance against powdery mildew (*Blumeria graminis* f. sp. *hordei*) enhance susceptibility to the rice blast fungus *Magnaporthe grisea*. *Molecular Plant-Microbe Interactions* **12**, 508-514.

Jasinski M., Ducos E., Martinoia E., Boutry M. (2003) The ATP-binding cassette transporters: Structure, function, and gene family comparison between rice and *Arabidopsis*. *Plant Physiology* **131**, 1169-1177.

John F.-N., Michael G. K J., Zhaohui W. (2009) Functional characterization of transcripts expressed in early-stage *Meloidogyne javanica* induced giant cells isolated by laser microdissection. *Molecular Plant Pathology* **10**, 237-248.

Kaiser C., Whiley A. W., Pegg K. G., Hargreaves P. A., Weinhart M. P. (1997) Determination of critical root concentrations of phosphate to control Phytophthora root rot in avocado. *Joint meeting of the Australian Avocado Grower's Federation Inc. and the New Zealand Avocado Grower's Association*, 155.

Kang J., Hwang J.-U., Lee M., Kim Y.-Y., Assmann S. M., Martinoia E., Lee Y. (2010) PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. *Proceedings of the National Academy of Sciences* **107**, 2355-2360.

© University of Pretoria

Kellam M. K., Coffey D. M. (1985) Quantitative comparison of the resistance to Phytophthora root rot in three avocado rootstocks. *Phytopathology* **75**, 230-234.

Kiba A., Maimbo M., Kanda A., Tomiyama H., Ohnishi K., Hikichi Y. (2007) Isolation and expression analysis of candidate genes related to *Ralstonia solanacearum*-tobacco interaction. *Plant Biotechnology* **24**, 409-416.

Köhne J. S. (2005) Selection of Avocado Scions and breeding of rootstocks in South Africa. In: *New Zealand and Australia Avocado Grower's Conference*, Tauranga, New Zealand.

Kumar J., Huckelhoven R., Beckhove U., Nagarajan S., Kogel K.-H. (2001) A Compromised Mlo pathway affects the response of barley to the necrotrophic fungus *Bipolaris sorokiniana* (teleomorph: *Cochliobolus sativus*) and its toxins. *Phytopathology* **91**, 127-133.

Kupke T., Hernández-Acosta P., Culiáñez-Macià F. A. (2003) 4′-phosphopantetheine and coenzyme A biosynthesis in plants. *Journal of Biological Chemistry* **278**, 38229-38237.

Lavi U., Akkaya M., Bhagwat A., Lahav E., Cregan P. B. (1994) Methodology of generation and characteristics of simple sequence repeat DNA markers in avocado (*Persea americana* M.). *Euphytica* **80**, 171-177.

Lavi U., Hillel J., Vainstein A., Lahav E., Sharon D. (1991) Application of DNA fingerprints for identification and genetic analysis of avocado. *Journal of the American Society for Horticultural Science* **116**, 1078-1081.

Liao Y., Zou H.-F., Wei W., Hao Y.-J., Tian A.-G., Huang J., Liu Y.-F., Zhang J.-S., Chen S.-Y. (2008) Soybean *GmbZIP44*, *GmbZIP62* and *GmbZIP78* genes function as negative regulator of ABA signaling and confer salt and freezing tolerance in transgenic *Arabidopsis*. *Planta* **228**, 225-240.

Liu J.-J., Sturrock R., Ekramoddoullah A. (2010) The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Reports* **29**, 419-436.

© University of Pretoria

Meyer E., Aglyamova G., Wang S., Buchanan-Carter J., Abrego D., Colbourne J., Willis B., Matz M. (2009) Sequencing and *de novo* analysis of a coral larval transcriptome using 454 GSFlx. *BMC Genomics* **10**, 219.

Mhameed S., Hillel J., Lahav E., Sharon D., Lavi U. (1995) Genetic association between DNA fingerprint fragments and loci controlling agriculturally important traits in avocado (*Persea americana* Mill.). *Euphytica* **84**, 81-87.

Mhameed S., Sharon D., Kaufman D., Lahav E., Hillel J., Degani C., Lavi U. (1997) Genetic relationships within avocado (*Persea americana* Mill) cultivars and between *Persea* species. *Theoretical and Applied Genetics* **94**, 279-286.

Mir G., Domenech J., Huguet G., Guo W.-J., Goldsbrough P., Atrian S., Molinas M. (2004) A plant type 2 metallothionein (MT) from cork tissue responds to oxidative stress. *Journal of Experimental Botany* **55**, 2483-2493.

Mittler R., Feng X., Cohen M. (1998) Post-ranscriptional suppression of cytosolic ascorbate peroxidase expression during pathogen-induced programmed cell death in tobacco. *The Plant Cell Online* **10**, 461-474.

Monteiro S., Barakat M., Picarra-Pereira M. A., Teixeira A. R., Ferreira R. B. (2003) Osmotin and thaumatin from grape: a putative general defence mechanism against pathogenic fungi. *Phytopathology* **93**, 1505-1512.

Moon D. H., Salvatierra G. R., Caldas D. G. G., Gallo de Carvalho M. C. C., Carneiro R. T., Franceschini L. M., Oda S., Labate C. A. (2007) Comparison of the expression profiles of susceptible and resistant *Eucalyptus grandis* exposed to *Puccinia psidii* Winter using SAGE. *Functional Plant Biology* **34**, 1010-1018.

Mousavi A., Hotta Y. (2005) Glycine-rich proteins. *Applied Biochemistry and Biotechnology* **120**, 169-174.

© University of Pretoria

Narayanan N., Grosic S., Tasma I., Grant D., Shoemaker R., Bhattacharyya M. (2009) Identification of candidate signaling genes including regulators of chromosome condensation 1 protein family differentially expressed in the soybean– *Phytophthora sojae* interaction. *Theoretical and Applied Genetics* **118**, 399-412.

Noctor G., Foyer C. H. (1998) Ascorbate and glutathione: keeping active oxygen under control. *Annual Review of Plant Physiology and Plant Molecular Biology* **49**, 249-279.

Opalski K. S., Schultheiss H., Kogel K.-H., Hückelhoven R. (2005) The receptor-like MLO protein and the RAC/ROP family G-protein RACB modulate actin reorganization in barley attacked by the biotrophic powdery mildew fungus *Blumeria graminis* f.sp. *hordei*. *The Plant Journal* **41**, 291-303.

Osman H., Vauthrin S., Mikes V., Milat M.-L., Panabières F., Marais A., Brunie S., Maume B., Ponchet M., Blein J.-P. (2001) Mediation of elicitin activity on tobacco is assumed by elicitin-sterol complexes. *Molecular Biology of the Cell* **12**, 2825-2834.

Park C.-J., Kim K.-J., Shin R., Park J. M., Shin Y.-C., Paek K.-H. (2004) Pathogenesis-related protein 10 isolated from hot pepper functions as a ribonuclease in an antiviral pathway. *The Plant Journal* **37**, 186-198.

Perrigault M., Tanguy A., Allam B. (2009) Identification and expression of differentially expressed genes in the hard clam, *Mercenaria mercenaria*, in response to quahog parasite unknown (QPX). *BMC Genomics* **10**, 377.

Piffanelli P., Zhou F., Casais C., Orme J., Jarosch B., Schaffrath U., Collins N. C., Panstruga R., Schulze-Lefert P. (2002) The barley MLO modulator of defence and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiology* **129**, 1076-1085.

Pnueli L., Hallak-Herr E., Rozenberg M., Cohen M., Goloubinoff P., Kaplan A., Mittler R. (2002) Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retama raetam*. *The Plant Journal* **31**, 319-330.

© University of Pretoria

Priest H. D., Fox S. E., Filichkin S. A., Mockler T. C. (2010) Utility of next-generation sequencing for analysis of horticultural crop transcriptomes. In: *International Symposium on Molecular Markers in Horticulture*, pp. 283-288. International Society for Horticultural Science Acta Horticulturae.

Rookes J. E., Wright M. L., Cahill D. M. (2008) Elucidation of defence responses and signalling pathways induced in *Arabidopsis thaliana* following challenge with *Phytophthora cinnamomi*. *Physiological and Molecular Plant Pathology* **72**, 151-161.

Sasabe M., Toyoda K., Shiraishi T., Inagaki Y., Ichinose Y. (2002) cDNA cloning and characterization of tobacco ABC transporter: *NtPDR1* is a novel elicitor-responsive gene. *Federation of European Biochemical Societies Letters* **518**, 164-168.

Sauter M., Rzewuski G., Marwedel T., Lorbiecke R. (2002) The novel ethylene-regulated gene *OsUsp1* from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *Journal of Experimental Botany* **53**, 2325-2331.

Scheibye-Alsing K., Hoffmann S., Frankel A., Jensen P., Stadler P. F., Mang Y., Tommerup N., Gilchrist M. J., Nygård A. B., Cirera S., Jørgensen C. B., Fredholm M., Gorodkin J. (2009) Sequence assembly. *Computational Biology and Chemistry* **33**, 121-136.

Schmelzer E. (2002) Cell polarization, a crucial process in fungal defence. *Trends in Plant Science* **7**, 411-415.

Schoch G. A., Nikov G. N., Alworth W. L., Werck-Reichhart D. (2002) Chemical inactivation of the cinnamate 4-hydroxylase allows for the accumulation of salicylic acid in elicited cells. *Plant Physiology* **130**, 1022-1031.

Schummer M., NG W.-I., Nelson P. S., Bumgarner R. E., Hood L. (1997) Inexpensive handheld device for the construction of high-density nucleic acid arrays. *BioTechniques* **23**, 1087-1092.

© University of Pretoria

Schütz I., Gus-Mayer S., Schmelzer E. (2006) Profilin and Rop GTPases are localized at infection sites of plant cells. *Protoplasma* **227**, 229-235.

Sharon D., Cregan P. B., Mhameed S., Kusharska M., Hillel J., Lahav E., Lavi U. (1997) An integrated genetic linkage map of avocado. *Theoretical and Applied Genetics* **95**, 911-921.

Sharon D., Hillel J., Mhameed S., Cregan P. B., Lavi U., Lahav E. (1998) Association between DNA markers and loci controlling avocado traits. *American Society for Horticultural Science* **123**, 1016-1022.

Showalter A., Kieliszewski M., Cheung A., Tierney M. (1996) Genes encoding cell wall proteins. *Plant Molecular Biology Reporter* **14**, 9-10.

Shumeta Z. (2010) Avocado production and marketing in Southwestern Ethiopia. *Trends in Agricultural Economics* **3**, 190-206.

Simon S. A., Zhai J., Nandety R. S., McCormick K. P., Zeng J., Mejia D., Meyers B. C. (2009) Short-read sequencing technologies for transcriptional analyses. *Annual Review of Plant Biology* **60**, 305-333.

Skøt L., Minchin F. R., Timms E., Fortune M. T., Webb K. J., Gordon A. J. (1996) Analysis of the two nodulins, sucrose synthase and ENOD2, in transgenic *Lotus* plants. *Plant and Soil* **186**, 99-106.

Stuible H.-P., Büttner D., Ehlting J., Hahlbrock K., Kombrink E. (2000) Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *Federation of European Biochemical Societies Letters* **467**, 117-122.

Sun H.-Q., Kwiatkowska K., Yin H. L. (1995) Actin monomer binding proteins. *Current Opinion in Cell Biology* **7**, 102-110.

© University of Pretoria

Thomas S. W., Glaring M. A., Rasmussen S. W., Kinane J. T., Oliver R. P. (2002) Transcript profiling in the barley mildew pathogen *Blumeria graminis* by serial analysis of gene expression (SAGE). *Molecular Plant-Microbe Interactions* **15**, 847-856.

Thomma B. P. H. J., Eggermont K., Penninckx I. A. M. A., Mauch-Mani B., Vogelsang R., Cammue B. P. A., Broekaert W. F. (1998) Separate jasmonate-dependent and salicylate-dependent defence-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. *Proceedings of the National Academy of Sciences* **95**, 15107-15111.

Tichtinsky G., Vanoosthuyse V., Cock J. M., Gaude T. (2003) Making inroads into plant receptor kinase signalling pathways. *Trends in Plant Science* **8**, 231-237.

Tilton G., Wedemeyer W., Browse J., Ohlrogge J. (2006) Plant coenzyme A biosynthesis: characterization of two pantothenate kinases from *Arabidopsis*. *Plant Molecular Biology* **61**, 629-642.

Tjaden J., Möhlmann T., Kampfenkel K., Neuhaus H. E., Henrichs G. (1998) Altered plastidic ATP/ADP-transporter activity influences potato (*Solanum tuberosum* L.) tuber morphology, yield and composition of tuber starch. *The Plant Journal* **16**, 531-540.

Uknes S., Mauch-Mani B., Moyer M., Potter S., Williams S., Dincher S., Chandler D., Slusarenko A., Ward E., Ryals J. (1992) Acquired resistance in *Arabidopsis*. *Plant Cell* **4**, 645-656.

van den Brûle S., Smart C. (2002) The plant PDR family of ABC transporters. *Planta* **216**, 95-106.

Vance C. P., Kirk T. K., Sherwood R. T. (1980) Lignification as a Mechanism of Disease Resistance. *Annual Review of Phytopathology* **18**, 259-288.

© University of Pretoria

Vega-Arreguin J., Ibarra-Laclette E., Jimenez-Moraila B., Martinez O., Vielle-Calzada J., Herrera-Estrella L., Herrera-Estrella A. (2009) Deep sampling of the *Palomero* maize transcriptome by a high throughput strategy of pyrosequencing. *BMC Genomics* **10**, 299.

Velculescu V. E., Zhang L., Vogelstein B., Kinzler K. W. (1995) Serial analysis of gene expression. *Science* **270**, 484-487.

Vera J. C., Wheat C. W., Fescemyer H. W., Frilander M. J., Crawford D. L., Hanski I., Marden J. H. (2008) Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17**, 1636-1647.

Wall P. K., Leebens-Mack J., Chanderbali A., Barakat A., Wolcott E., Liang H., Landherr L., Tomsho L., Hu Y., Carlson J., Ma H., Schuster S., Soltis D., Soltis P., Altman N., dePamphilis C. (2009) Comparison of next generation sequencing technologies for transcriptome characterization. *BMC Genomics* **10**, 347.

Wang X., Tang C., Deng L., Cai G., Liu X., Liu B., Han Q., Buchenauer H., Wei G., Han D., Huang L., Kang Z. (2010) Characterization of a pathogenesis-related thaumatin-like protein gene TaPR5 from wheat induced by stripe rust fungus. *Physiologia Plantarum* **139**, 27-38.

Weber A. P., Weber K. L., Carr K., Wilkerson C., Ohlrogge J. B. (2007) Sampling the *Arabidopsis* transcriptome with massively parallel pyrosequencing. *Plant Physiology* **144**, 32 - 42.

Werck-Reichhart D., Feyereisen R. (2000) Cytochromes P450: a success story. *Genome Biology* **1**, 30031-30039.

Zentmyer G. A. (1984) Avocado diseases. *Tropical Pest Managment* **30**, 388-400.

Zentmyer G. A. (1955) Diseases of the Avocado. *California Avocado Society Yearbook* **39**, 44-58.

© University of Pretoria

2.6 Figures & Tables

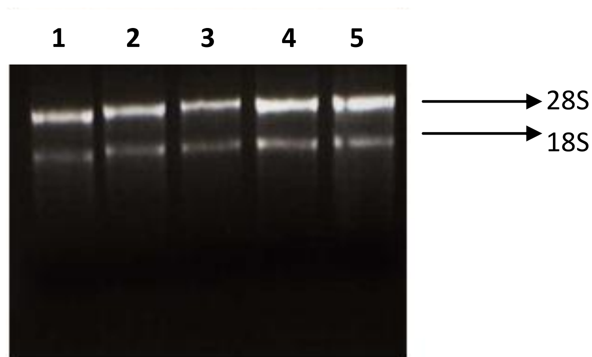


Figure 1. RNA extraction of *Phytophthora cinnamomi* infected tolerant R0.09 avocado roots. Lanes 1-5; RNA extracted from roots of avocado rootstock, R0.09, at various time points with two distinct 28S and 18S rRNA bands. Lane 1: 0 hour RNA isolation, lane 2: 6 hour RNA isolation, lane 3:12 hour RNA isolation, lane 4:24 hour RNA isolation and lane 5: 48 hour RNA isolation.

© University of Pretoria

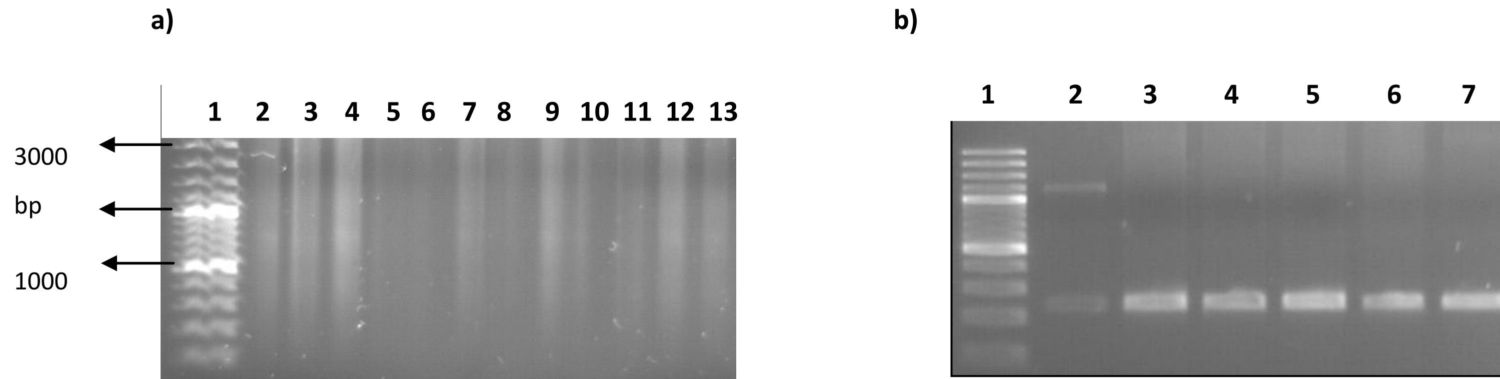


Figure 2. Product of cDNA synthesis from mRNA of *Phytophthora cinnamomi* infected R0.09 avocado roots. Lane 1: DNA molecular marker, lane 2-3: 0 hr cDNA, lanes 4-7: infected library 1cDNA, lanes 8-12 cDNA synthesis of infected library 2 of R0.09 rootstock and lane 13: positive control. A smear between 400bp and 3kb indicate the successful synthesis of cDNA.

b) Testing of cDNA for DNA contamination. Lane 1: molecular marker 100bp, lane 2: Avocado DNA, lane3- 7: 0 hr, 6hr, 12hr, 24hr 48hr cDNA synthesis samples respectively. Amplification of genomic DNA results in a ~1200 bp (lane 2) band and the cDNA equivalent is a band of ~240 bp. All samples were negative for the 1200 bp band but positive for the smaller cDNA band. This indicated no DNA contamination.

© University of Pretoria

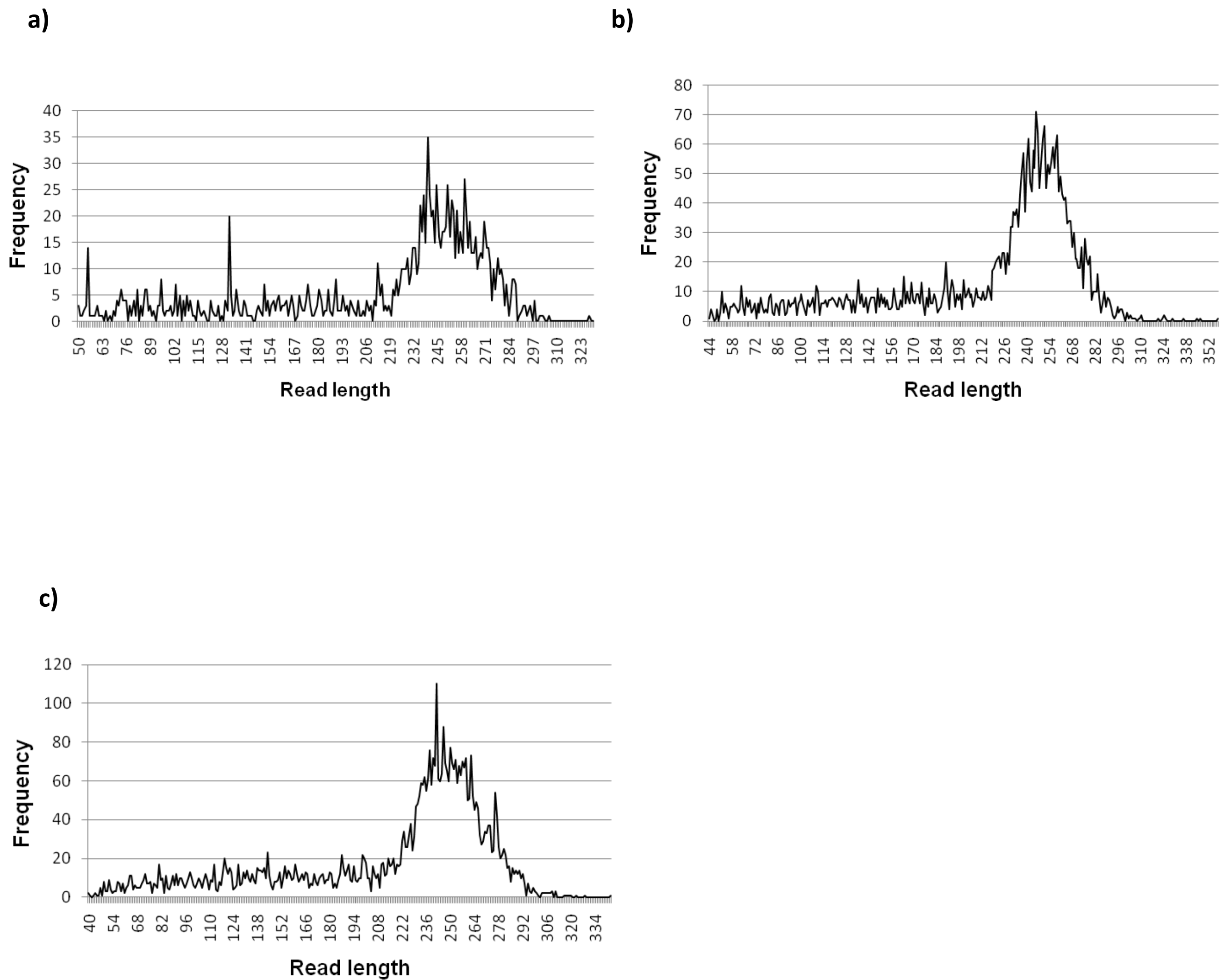
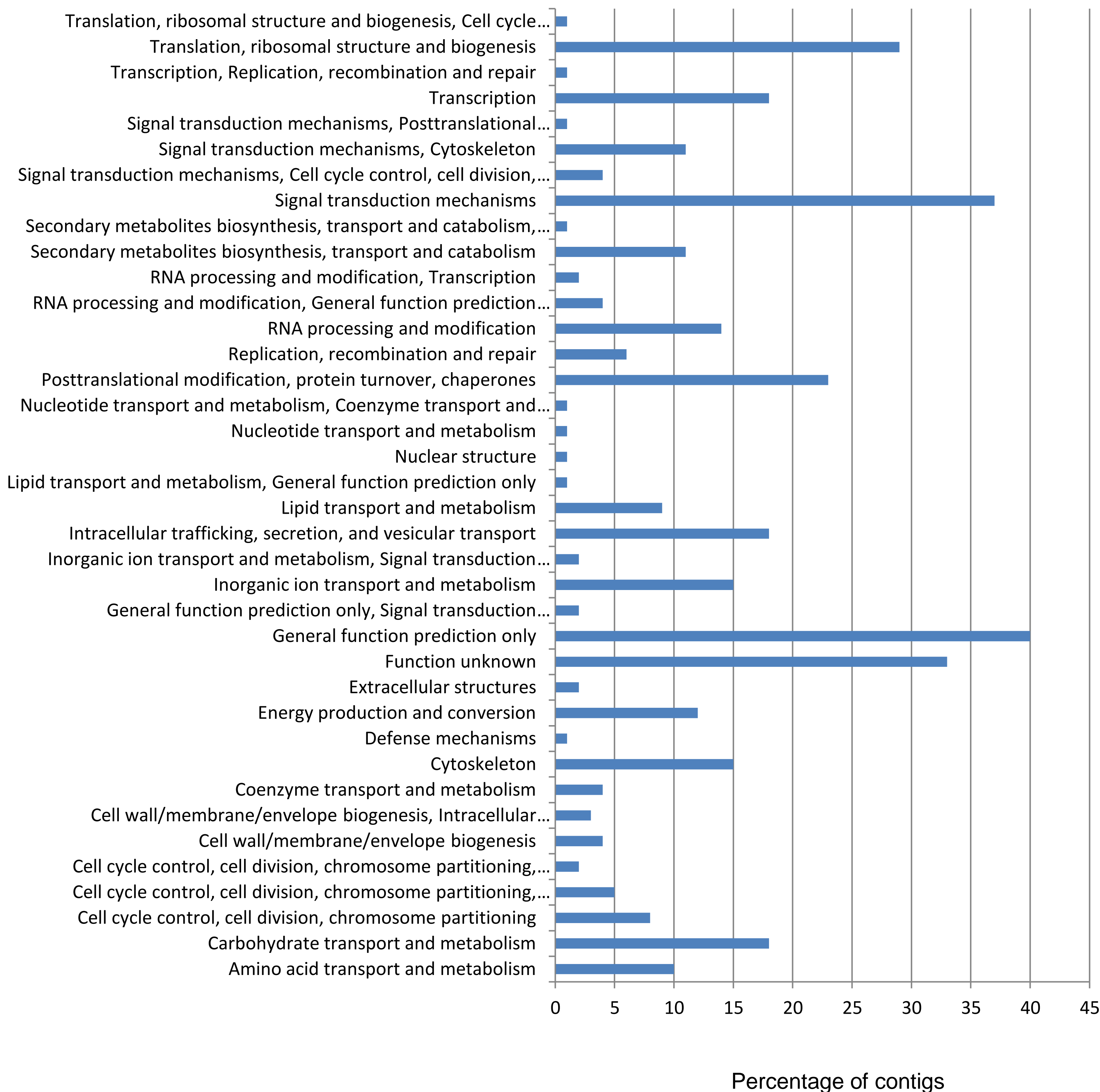


Figure 3. Read length distributions of uninfected, infected library 1 and infected library 2 from *Phytophthora cinnamomi* infected tolerant R0.09 avocado roots (a) Uninfected library has reads with the highest frequency at around 245 bp. **(b)** Infected library 1 reads have the highest frequency at around 252 bp. **(c)** Infected library 2 reads have the highest frequency at around 240 bp.

© University of Pretoria



© University of Pretoria

Figure 4. KOG classifications of avocado transcripts identified in all three cDNA libraries (uninfected, infected libraries 1 and 2) of *Phytophthora cinnamomi* infected R0.09 avocado roots. The KOG database is essentially a eukaryotic version of the COG database (Clusters of Orthologous Groups). The largest number of classifications belongs to the classes of general function prediction and unknown functions; this is followed by signal transduction mechanisms and translational mechanisms.

© University of Pretoria

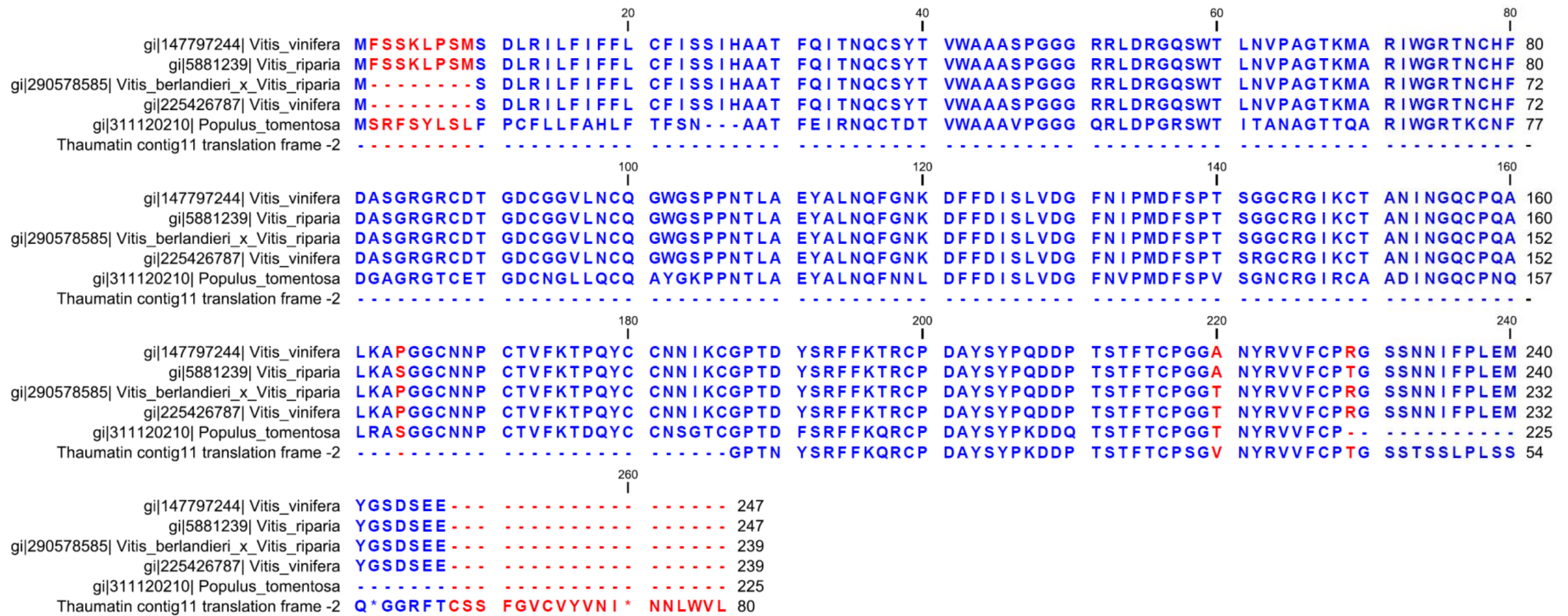


Figure 5. Thaumatin protein from infected library 2 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The thaumatin EST amino acid sequence showed close homology to genes from *Vitis vinifera, V. riparia, V. riparia crossed* with *V. berlandieri* and *Populus tomentosa.* The sequence alignment shows conserved amino acid sequence in blue with the non-conserved regions in red.

© University of Pretoria

Figure 6. Thaumatin-like protein from infected library 2 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The thaumatin-like EST amino acid sequence showed close homology to genes from *Vitis vinifera, Arabidopsis thaliana, Oryza sativa* and *Nicotiana tabacum.* The sequence alignment shows highly conserved amino acid sequence in blue with partially conserved regions in black and non-conserved regions in red.

© University of Pretoria

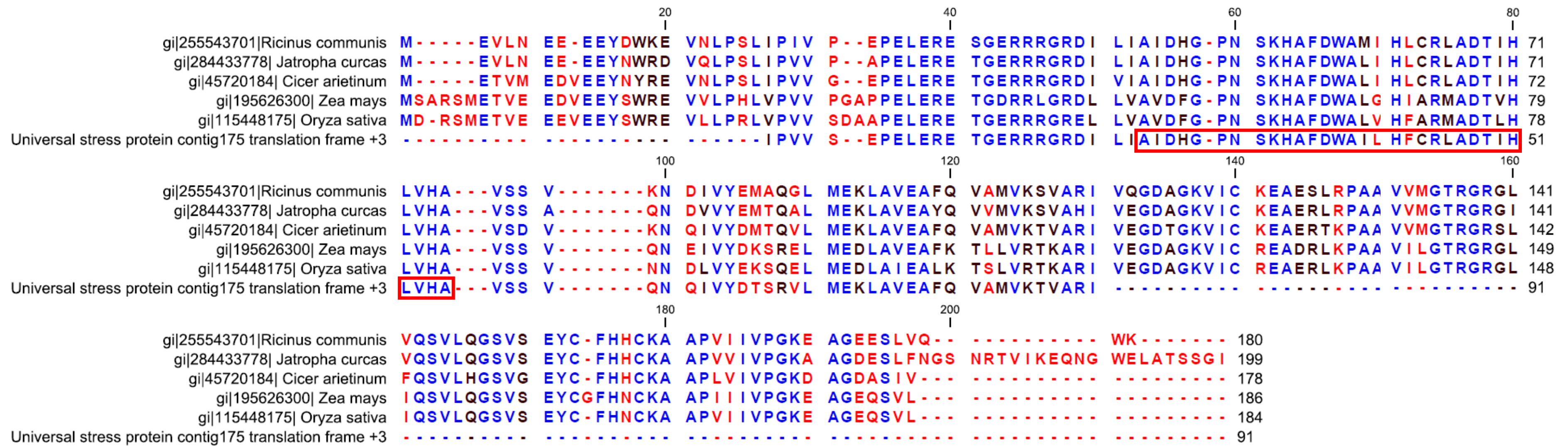


Figure 7. Universal stress protein from the infected library 2 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The universal stress protein EST amino acid sequence showed close homology to genes from *Ricinus communis, Jatropha curcas, Cicer arietinum, Oryza sativa* and *Zea mays.* The sequence alignment shows highly conserved amino acid sequence in blue with partially conserved regions in black and non-conserved regions in red. Red box indicates the region containing the conserved residues specific to the universal stress protein ligand binding domain.

© University of Pretoria

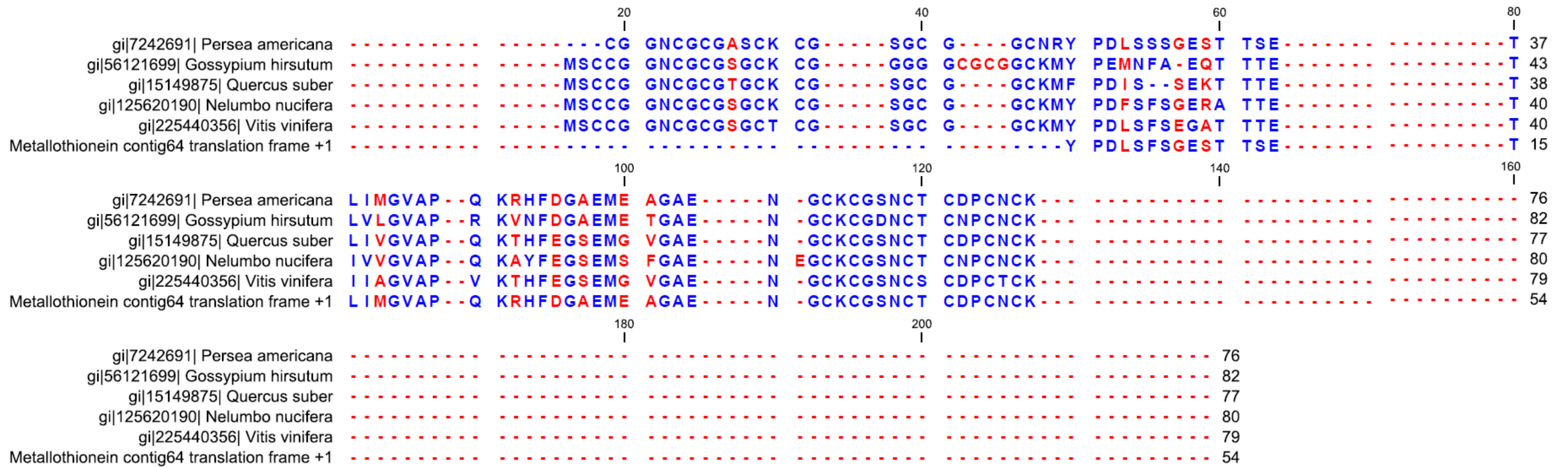


Figure 8. Metallothionein protein from the infected library 2 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The metallothionein EST amino acid sequence showed close homology to genes from *Persea americana, Vitis vinifera, Gossypium hirsutum, Quercus suber* and *Nelumbo nucifera.* The sequence alignment shows highly conserved amino acid sequence in blue with non-conserved regions in red.

© University of Pretoria

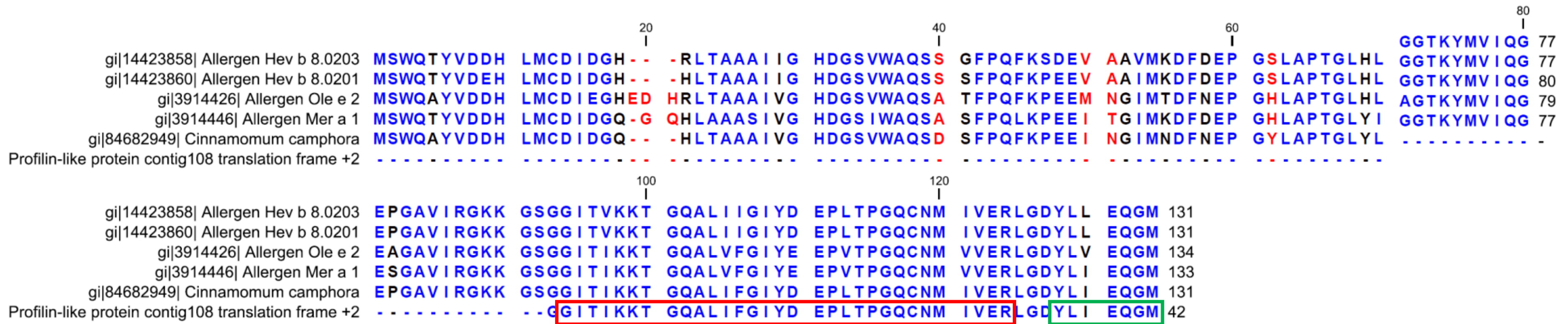


Figure 9. Profilin-like protein from infected library 2 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The profilin-like EST amino acid sequence showed close homology to allergen genes from *Hevea brasiliensis* and *Mercurialis annua* and to *Cinnamomum camphora.* The sequence alignment shows highly conserved amino acid sequence in blue with partially conserved regions in black and non-conserved regions in red. Red box is the actin interaction site while the green box is the poly-proline binding site.

© University of Pretoria

```
                                                                    20                                  40                                  60                                  80
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera M--------- ----AG---- --DSSSSTSA SRELDQTPTW AVAGICAVMI IISIVLEKVL HRVGKWFTER RKRALFEALE 61
                  gi|317457431| Malus toringoides MVPFSLLGVS LCLVAGLLDT AMGAGKTTSY SRELDRTPTW AVAGVCAVII IISLVLEKVL HKVGIWLTDR HKRALFESLE 80
                  gi|255583426| Ricinus communis MA-------- ---------- ---ASDTTAT QRKLDQTPTW AVASVCAVMI IISILLEKGL HKFGTWLTHK HKRALFEALE 59
                  gi|255539290| Ricinus communis M--------- ---------- ---AGGETPG ARELDQTPTW AVAAVCAIIV LLSIILEKFL HWTGEWFQER HKRALYEALE 58
                         gi|162461110| Zea mays M--------- ---------- --GGGGGGGN SRELDQTPTW AVASVCGVIV LISILLEKGL HHVGEFFSHR KKKAMVEALE 59
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -
                                                                   100                                 120                                 140                                 160
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera KVKAELMVLG FISLLLTFGQ NFIVKICIPE KAADTMLPCP YNGE----KD SSSET----E S-RRRLL--- ---------- 119
                  gi|317457431| Malus toringoides KVKAELMILG FISLILTFGQ SYIAKICIPL NVADTMLPCR VDDK----E- -KEES----T S-DRRLL--- ---------- 136
                  gi|255583426| Ricinus communis KVKAELMVLG FISLLLTFGQ TYITRICVSQ KVADTMLPCK GDAV----LD TEEST----E EHRRRLL--- ---------- 118
                  gi|255539290| Ricinus communis KVKGELMVLG FISLLLTFSQ NFIARICIRE KFANTMLPCP QKGK----HH EPDSDPFKEE DHHRRLL--- ---------- 121
                         gi|162461110| Zea mays KVKAELMVLG FISLLLVFGQ NYIIKVCISN HAANTMLPCK LEAAAVEGKD GHGKEAAAVV AGKKKVAVAV PGKKKKKAAA 139
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- -
                                                                   180                                 200                                 220                                 240
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera ---------- W------YNH RLLAAATYSS SCKEGYEPII SVNGLHQLHI LIFFLAVFHV LYSAITMMLG RLKIRGWKQW 183
                  gi|317457431| Malus toringoides ---------- W------YER RSLAAAS-EY KCKTGYEPLI TVNGLHQLHI LIFFLAAFHV LYSLLTMLLG RLKIRGWKHW 199
                  gi|255583426| Ricinus communis ---------- W------FER RFLAGAETTT LCKEGHEPLI SADGLHQLHI LIFFLAVFHV VYSLTTMTLG RLKIRGWKAW 182
                  gi|255539290| Ricinus communis ---------- L------YDR RFLAADSSGA GCKAGHVPLI SVNGLHQLHI FIFFLAVFHV IYSAITMMLG RLKIRGWKEW 185
                         gi|162461110| Zea mays AADHLGGVVD WPPPYYAHNA RMLAEASMAT KCPEGKVPLI SINALHQLHI FIFFLAVFHV SYSAITMALG RAKIRAWKEW 219
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- ---------G RLKIRGWKEW 11
                                                                   260                                 280                                 300                                 320
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera EEETSTHDYE FSNDAARFRL THETSFVRAH TSFWTRIPFF FYVGCFFRQF FRSVSRSDYL TLRNGFITVH LAPGSKFNFQ 263
                  gi|317457431| Malus toringoides EAETSSDDYE FSNDPSRFRL THQISFVRAH TSFWTKIPFF FYFGCFFRQF FRSVSKADYL TLRNGFITVH LGAGSKFNFQ 279
                  gi|255583426| Ricinus communis EEETSTHDYE FSNDPSRFRL THETSFVRSH TSFWTRIPFF FYVGCFFRQF FRSVSKADYL TLRNGFITVH LAPGSKFNFQ 262
                  gi|255539290| Ricinus communis ERE-SSKDHD ALNDPTRFRL THETSFVRDH TSFWTKMPAL HYVVCFFRQF FRSVTKADYL AMRHGFISVH LAPGSKFDFQ 264
                         gi|162461110| Zea mays EKEAAGQDYE FSHDPTRFRF THETSFVRQH MNVLNKFPAS FYISNFFRQF FRSVRQADYC ALRHSFVNVH LAPGSKFDFQ 299
MLO transmembrane protein contig88 translation frame +1 EKETRSIDYE FSNDPTRFRF AHETSFVRQH ASVWNRIPVL FYVVSFFRQF FWSVRKADYL AMRHGFIVVH LA-------- 83
                                                                   340                                 360                                 380                                 400
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera KYIKRSLEDD FKVVVGISPV LWASFVIFLL LNVSGWQAMF WASIIPLVIV FAVGTKLQAV LTKMALEITE RHAVVQGIPL 343
                  gi|317457431| Malus toringoides KYIKRSLEDD FKVVVGVSPL LWASFVIFLL LNVKGWQALF WASLIPLVII LLVGTKLQAI LTKMAIEIAE RHAVVQGIPL 359
                  gi|255583426| Ricinus communis KYIKRSLEDD FKYVVGISPI LWASFVLFLL INVKGWQALF WASLIPVIII LAVGTELQSV LTKMALEIAE RHAVVQGMPL 342
                  gi|255539290| Ricinus communis KYIKRSLEDD FKVVVGISPL LWASVVVFLL FNVHGWQAIY WVSVLPIFVI LAVGTKLQAI ITQMAVEIQE RHAVVQGIPL 344
                         gi|162461110| Zea mays KYIKRSLEDD FKVIVGISPP LWASALIFLF LNVNGWHTML WISIMPVVII LSVGTKLQGI ICRMAIDITE RHAVIQGIPM 379
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 83
                                                                   420                                 440                                 460                                 480
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera VQGSDKYFWF SWPQLVLHLI HFVLFQNAFQ ITYFLWIWYS FGLKSCFHNN FKLVIIKIAL GVGVLFLCSY ITLPLYALIT 423
                  gi|317457431| Malus toringoides VQGSDKYFWF GRPQLILNLI HFALFQNAFQ IIYFFWIWYS FGLKSCFHAN FKLAIAKVIL GVGVLCLCSY ITLSLYALVT 439
                  gi|255583426| Ricinus communis VQGSDKYFWF GRPQLVLYLI HFALFQNAFQ ITYFLWIWYE FGLKSCFHAN FTLAIVKVSL GL-------- ---------- 404
                  gi|255539290| Ricinus communis VQVSDKHFWF SWPQLVLFLI HFVLFQNAFE LTYFFWIWYE FGLKSCFHED INLIYLRVAL G--------- ---------- 405
                         gi|162461110| Zea mays VQVSDSYFWF ARPTFVLFLI HFTLFQNGFQ IIYFLWILYE YGMDSCFNDS EEFVFARLCL GVVVQVLCSY VTLPLYALVS 459
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 83
                                                                   500                                 520                                 540                                 560
                                                                     |                                   |                                   |                                   |
                     gi|225438432| Vitis vinifera QMGSNMKRSV FDEQTSKALK KWHMAVKKR- HGGKAGRSST HQTLGGSPTA SMASTVQMPG SGHALHRFKT TGHSTRTFTY 502
                  gi|317457431| Malus toringoides QMGSHMKRAI FDEQTSKALK KWHMAVKKKT HGGK----SP TRTLGGE--S STISTMRSST SGHTLHRFKT TGHSTRSAAF 513
                  gi|255583426| Ricinus communis ---------- ---------- ---------- ---------- ---------- SPMSTVHSPP DVHC------ --------TV 420
                  gi|255539290| Ricinus communis ---------- ---------- ---------- ---------- ---------- PSSSPAQSPS HMHSHPR--- ---------- 422
                         gi|162461110| Zea mays QMGSTMKQSI FDEQTSKALK NWRAGAKKKA PTGGSKHG-- ----GGGSPT AGGSPTKADG D--------- ---------- 514
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- ---------- ---------- 83
                                                                   580                                 600                                 620
                                                                     |                                   |                                   |
                     gi|225438432| Vitis vinifera EDHEVSDYEG EPESPTSSTT NLIVRVDDSE PETEPVVSHP EHDTSGEIEF SFVKPAPQKG --P 563
                  gi|317457431| Malus toringoides EDHETSDPET DPQSP-SSTT HLIVRVDQIE QQTELNEPHD GEQTNIPDDF SFIPPAPGKE --T 573
                  gi|255583426| Ricinus communis QNHWALSP-- ---------- -LI------- ---------- ---------- -FLG-GPGH- --V 438
                  gi|255539290| Ricinus communis -DHELTDVEA DGADPHQ-TA NIMAIVDLDG DQEKGSHPY- ---------- -----GQQHD LLS 467
                         gi|162461110| Zea mays ---------- ---------- ---------- ---------- ---------- ---------- --A 515
MLO transmembrane protein contig88 translation frame +1 ---------- ---------- ---------- ---------- ---------- ---------- --- 83
```


© University of Pretoria

Figure 10. MLO transmembrane protein from infected library 1 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The MLO transmembrane protein EST amino acid sequence showed close homology to genes from *Vitis vinifera, Ricinus communis, Malus toringoides* and *Zea mays.* The sequence alignment shows highly conserved amino acid sequence in blue with partially conserved regions in black and non-conserved regions in red.

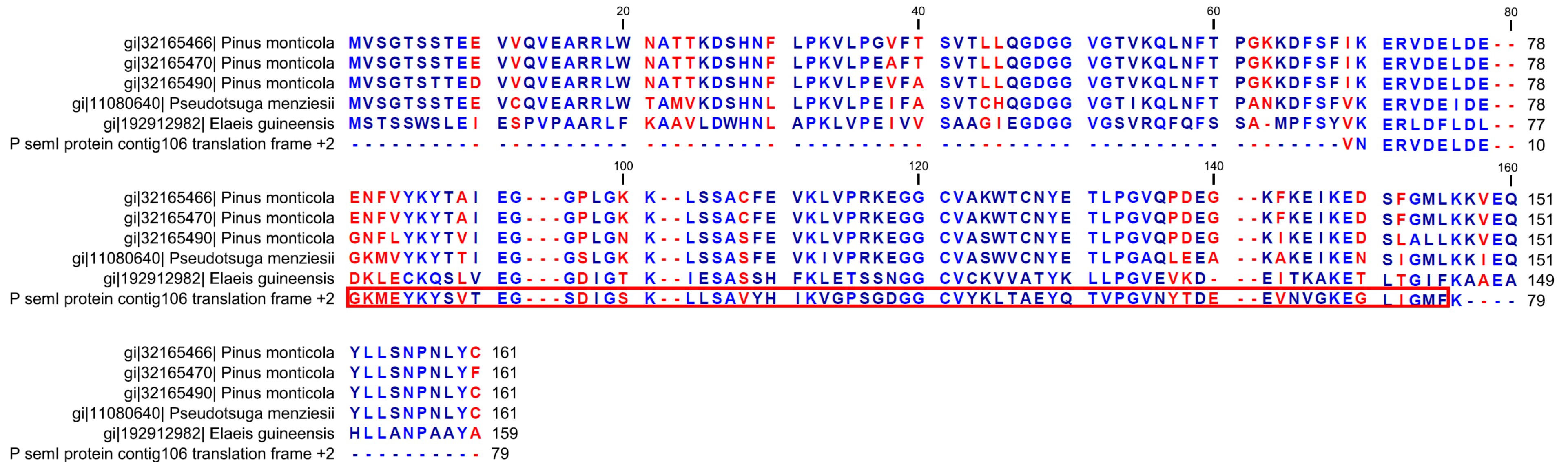


Figure 11. PsemI PR 10 protein from infected library 1 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The PR 10 protein EST amino acid sequence showed close homology to genes from *Pinus monticola, Pseudotsuga menziesii* and *Elaeis guineensis.* The sequence alignment shows highly conserved amino acid sequences in blue with partially conserved regions in darker blue and non-conserved regions in red. The red box highlights the 22 of 37 residues that comprise the characteristic hydrophobic ligand binding region of this family.

© University of Pretoria

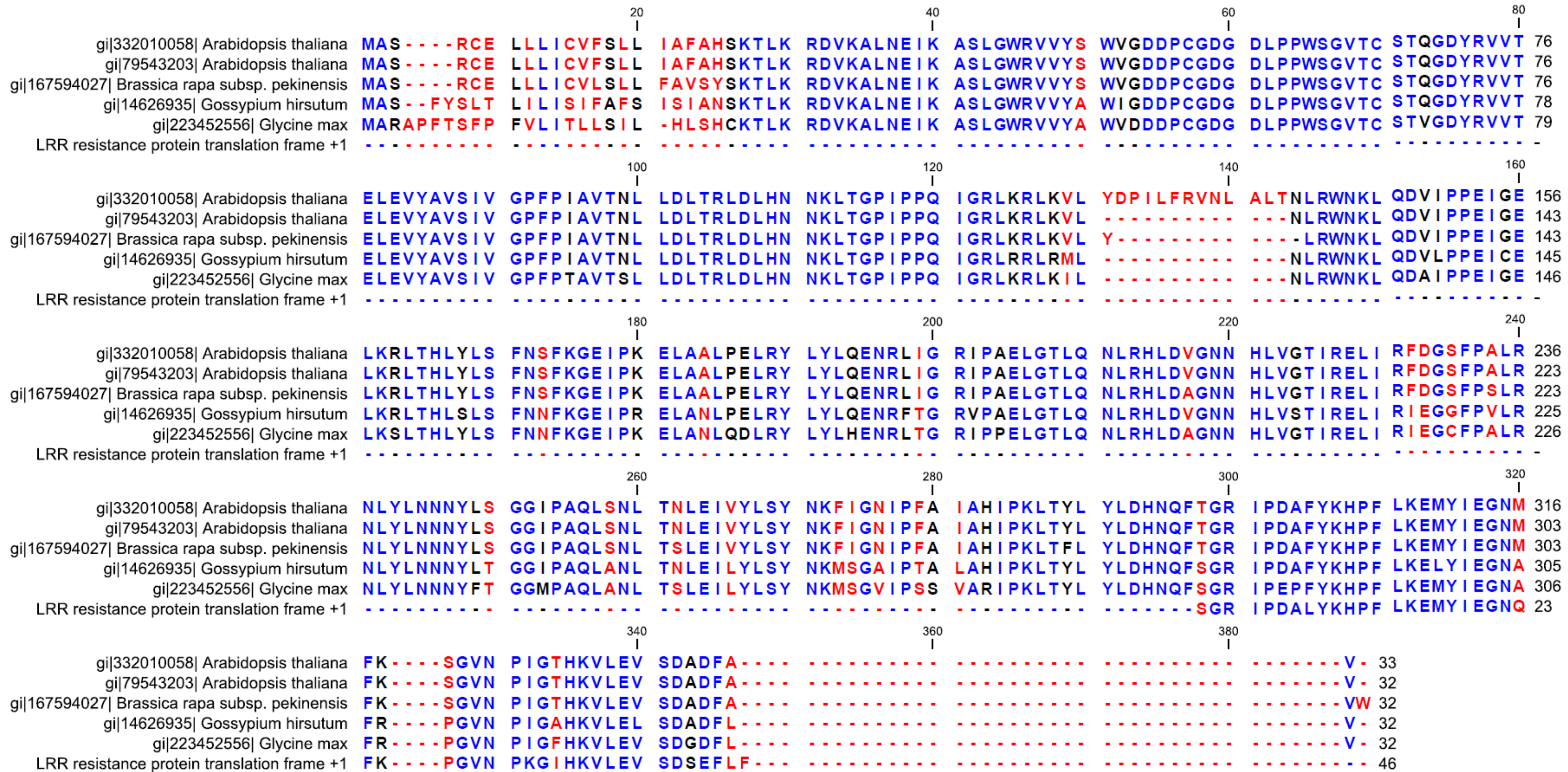


Figure 12. LRR resistance protein from the infected library 1 of *Phytophthora cinnamomi* infected avocado roots aligned to top five BLAST homologies. The Leucine-rich repeat protein EST amino acid sequence showed close homology to genes from *Arabidopsis thaliana, Gossypium hirsutum, Glycine max* and a *Brassica rapa* subspecies. The sequence alignment shows highly conserved amino acid sequences in blue with partially conserved regions in black and non-conserved regions in red.

© University of Pretoria

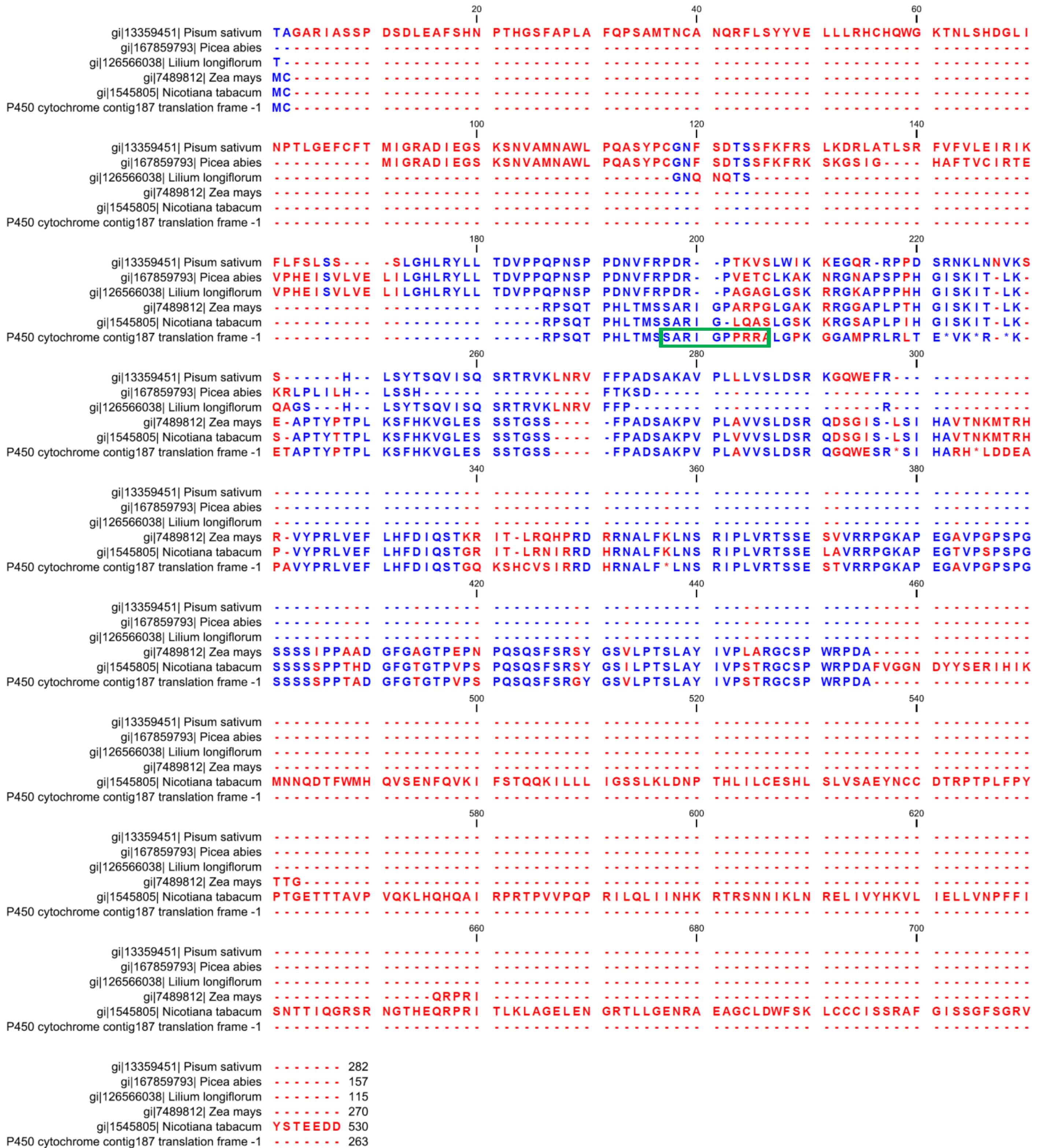


Figure 13. Cytochrome P450 from infected library 1 of *Phytophthora cinnamomi* infected avocado root aligned to top five BLAST homologies. The cytochrome P450 EST amino acid sequence showed close homology to genes from *Pisum sativum, Picea abies, Lillium longiform, Nicotiana tabacum* and *Zea mays.* The sequence alignment shows highly conserved amino acid sequences in blue with non-

© University of Pretoria

conserved regions in red. The green box identifies the area containing four of 19 conserved amino acid residues.

Table 1. Pyrosequencing tags of cDNA libraries for uninfected and *Phytophthora cinnamomi* infected roots.

cDNA library	Sequencing tag (5'-3')
Uninfected	CGTGTCTCTA
Infected library 1	CTCGCGTGTC
Infected library 2	TAGTATCAGC

Table 2. Excerpts of newblermetric reports from the uninfected, infected library 1 and infected library 2 from *Phytophthora cinnamomi* infected R0.09 avocado roots. The total number of reads and contigs are shown to illustrate the efficiency of the pyrosequencing run. Infected library 2 generated the largest amount of data with the uninfected library generating the least sequence data.

Libraries:	Uninfected	Infected library 1	Infected library 2
Total Reads	1407	3582	4961
Total Bases	288885	737254	1017064
Number of Contigs	43	139	189

© University of Pretoria

Table 3. List of total contigs generated from pyrosequencing of *Phytophthora cinnamomi* infected R0.09 avocado root. The contigs were annotated according to NR database. Of the 371 contigs, 192 contigs had positive gene identities, while 123 contigs were identified only as hypothetical proteins. There were 43 contigs that could not be aligned to any sequence in the NR database based on nucleotide or amino acid sequence.

Contig	Identity	E Value	Sequence ID	% Identity	Species
Uninfected library					
Positive gene identities					
contig00001	26S ribosomal RNA gene	3E-60	gi|1545805	100	*Magnolia denudata*
contig00002	T6D22.2	5E-12	gi|8778823	93	*Arabidopsis thaliana*
contig00003	18S Ribosomal RNA	2E-34	gi|50312723	73	*Kluyveromyces lactis*
contig00006	Late embryogenesis abundant protein Lea5	0.046	gi|22653808	59	*Citrus sinensis*
contig00011	18S Ribosomal RNA	3E-28	gi|147826550	95	*Vitis vinifera*
contig00014	Chain A, Solution Structrue Of The Zf-An1 Domain from At2g36320 Protein	8E-20	gi|159163297	70	*Arabidopsis thaliana*
contig00017	60S ribosomal protein L36 homolog	0.016	gi|19979631	95	*Allium cepa*
contig00018	NADH:cytochrome b5 reductase	4E-39	gi|55979111	91	*Vernicia fordii*
contig00024	CHR4/MI-2-LIKE (chromatin remodeling 4); ATP binding / DNA binding / chromatin binding / helicase	4E-19	gi|145358870	58	*Arabidopsis thaliana*
contig00027	LRX1 (LEUCINE-RICH REPEAT/EXTENSIN 1); protein binding / structural constituent of cell wall	0.007	gi|15221178	37	*Arabidopsis thaliana*
contig00028	ZIP family metal transporter	2E-12	gi|82581289	55	*Chengiopanax sciadophylloides*
contig00031	protein kinase family protein	3E-15	gi|30696480	86	*Arabidopsis thaliana*
contig00037	RNA and export factor-binding protein	3E-09	gi|145334661	52	*Arabidopsis thaliana*
contig00038	oil palm polygalacturonase allergen PEST472	4E-31	gi|34223513	83	*Elaeis guineensis*
contig00039	putative transcription factor	2E-25	gi|14582465	89	*Vitis vinifera*
contig00043	putative extensin	2E-08	gi|148807133	63	*Prunus dulcis*

© University of Pretoria

Defence-related genes					
contig00033	metallothionein-like protein type 2	6E-41	gi|7242691	94	*Persea americana*
Hypothetical proteins					
contig00004	hypothetical protein MGG_06317	3.3	gi|145608858	31	*Magnaporthe grisea 70-15*
contig00005	conserved hypothetical protein	9E-29	gi|90419459	82	*Aurantimonas sp. SI85-9A1*
contig00009	hypothetical protein	2E-35	gi|147858131	93	*Vitis vinifera*
contig00010	hypothetical protein	4.3	gi|147812476	35	*Vitis vinifera*
contig00015	hypothetical protein	1E-29	gi|147822174	79	*Vitis vinifera*
contig00021	hypothetical protein	1E-28	gi|147798057	74	*Vitis vinifera*
contig00025	hypothetical protein	0.016	gi|147815590	48	*Vitis vinifera*
contig00029	hypothetical protein	0.0006	gi|147781769	45	*Vitis vinifera*
contig00030	hypothetical protein OsI_013039	2E-08	gi|125545667	45	*Oryza sativa*
contig00032	hypothetical protein	2E-37	gi|147775208	92	*Vitis vinifera*
contig00034	hypothetical protein	5E-11	gi|147841226	49	*Vitis vinifera*
contig00035	hypothetical protein	1E-20	gi|147769159	68	*Vitis vinifera*
contig00036	hypothetical protein	5E-31	gi|147834261	88	*Vitis vinifera*
contig00042	hypothetical protein	0.0002	gi|147792026	51	*Vitis vinifera*
Genes out of plant kingdom					
contig00008	secreted antigen 1	9.6	gi|118442730	31	*Babesia gibsoni*
contig00013	Structure Of Farnesyl Pyrophosphate Synthetase Chain	9.7	gi|1942048	35	Unknown
contig00016	Na(+)-translocating NADH-quinone reductase subunit B	5.5	gi|149177061	47	*Planctomyces maris* DSM 8797
contig00020	hypothetical protein COSY_0553	7.2	gi|148244700	52	*Candidatus vesicomyosocius*
contig00023	hypothetical protein PARMER_02310	5.5	gi|154492673	30	*Parabacteroides merdae*
contig00026	polyprotein	5.6	gi|156447514	36	*Banana bract mosaic* virus

© University of Pretoria

Unknown proteins					
contig00019	unknown protein	0.0008	gi|18394157	46	*Arabidopsis thaliana*
Unannotated contigs					
contig00007					
contig00012					
contig00022					
contig00040					
contig00041					
Infected library 1					
Positive gene identities					
contig0018	dormancy-associated protein/auxin-repressed protein	1E-09	gi|148872940	82	*Glycyrrhiza uralensis*
contig0024	Elongation factor 2 (EF-2) elongation factor 2	0.003	gi|6015065	95	*Beta vulgaris* subsp. *vulgaris*
contig0025	ENOD2f	1E-14	gi|10198184	45	*Maackia amurensis*
contig0030	Major latex allergen Hev b 5 latex allergen pI 3.5 latex allergen	0.079	gi|7387766	38	*Hevea brasiliensis*
contig0034	alpha-amylase inhibitor-like precursor	9.8	gi|56237708	31	*Phaseolus maculatus*
contig0036	putative polyprotein	0.001	gi|54291863	46	*Oryza sativa*
contig0037	pentatricopeptide (PPR) repeat-containing protein	9E-08	gi|15218705	63	*Arabidopsis thaliana*
contig0038	caffeoyl-CoA 3-O-methyltransferase, putative	2E-26	gi|18416703	61	*Arabidopsis thaliana*
contig0047	bZIP transcription factor bZIP80	3E-30	gi|113367194	76	*Glycine max*
contig0060	pantothenate kinase family protein	1E-32	gi|18418688	77	*Arabidopsis thaliana*
contig0064	putative glycine rich protein	4E-36	gi|20152613	77	*Rumex obtusifolius*
contig0072	Putative small nuclear ribonucleoprotein G	0.0000005	gi|14192871	90	*Oryza sativa*
contig0076	metallothionein-like protein type 2	6E-41	gi|7242691	94	*Persea americana*
contig0081	putative AP3-complex beta-3A adaptin subunit	0.5	gi|16604671	47	*Arabidopsis thaliana*
contig0083	putative DnaJ protein	0.002	gi|145688390	100	*Camellia sinensis*
contig0086	glycyl-tRNA synthetase	9E-27	gi|110740563	80	*Arabidopsis thaliana*
contig0089	proline-rich extensin, putative	2E-10	gi|15222668	46	*Arabidopsis thaliana*

© University of Pretoria

contig0091	putative plastidic adenylate transporter	2E-33	gi|61651614	83	*Mesembryanthemum crystallinum*
contig0092	ATP synthase epsilon chain, mitochondrial	5E-28	gi|15217996	79	*Arabidopsis thaliana*
contig0096	PLA IIB/PLP6 (Patatin-like protein 6); nutrient reservoir	1E-10	gi|15225054	43	*Arabidopsis thaliana*
contig0098	tRNA synthetase class II (G, H, P and S) family protein	1E-08	gi|15228692	60	*Arabidopsis thaliana*
contig0104	sugar binding / transferase, transferring glycosyl groups	1E-29	gi|145335796	83	*Arabidopsis thaliana*
contig0111	70 kDa peptidyl-prolyl isomerase (Peptidyl-prolyl cis-trans isomerase) (PPIase)	4E-32	gi|3023751	86	*Triticum aestivum*
contig0113	ACT1	1E-14	gi|149938964	100	*Actinidia deliciosa*
contig0115	60S ribosomal protein L10 (QM protein homolog)	1E-08	gi|18203445	86	*Vitis riparia*
contig0120	Auxin-induced protein PCNT115 auxin-induced protein	8E-36	gi|728744	92	*Nicotiana tabacum*
contig0131	protein binding / zinc ion binding	1E-28	gi|145339866	81	*Arabidopsis thaliana*
contig0136	60S ribosomal protein L10 60S ribosomal protein L10	7E-27	gi|18203270	100	*Euphorbia esula*
contig0137	extensin	9E-14	gi|1486263	66	*Catharanthus roseus*
contig0139	extensin-like protein	5E-09	gi|791148	75	*Vigna unguiculata*
Defence-related genes					
contig0020	leucine-rich repeat resistance protein-like protein	8E-15	gi|14626935	80	*Gossypium hirsutum*
contig0070	4-coumarate-CoA ligase-like protein	5E-26	gi|29893225	69	*Arabidopsis thaliana*
contig0088	seven transmembrane protein Mlo6	1E-28	gi|13784985	69	*Zea mays*
contig0106	pathogenesis-related protein PsemI	5E-14	gi|11080640	46	*Pseudotsuga menziesii*
contig0109	drought-induced protein	8E-13	gi|14994231	68	*Retama raetam*
Hypothetical proteins					
contig0001	hypothetical protein	0.003	gi|147777692	51	*Vitis vinifera*
contig0002	hypothetical protein FG01701.1	3.3	gi|46109638	45	*Gibberella zeae PH-1*
contig0003	hypothetical protein	1E-49	gi|147835464	69	*Vitis vinifera*
contig0004	hypothetical protein PFE0570w	4.3	gi|124506149	33	*Plasmodium falciparum 3D7*

© University of Pretoria

contig0010	hypothetical protein DDBDRAFT_0202116	1.1	gi|66827755	34	*Dictyostelium discoideum AX4*
contig0013	hypothetical protein	0.003	gi|147800856	100	*Vitis vinifera*
contig0014	hypothetical protein	1E-15	gi|147865432	78	*Vitis vinifera*
contig0015	hypothetical protein	1E-24	gi|147770332	81	*Vitis vinifera*
contig0023	Os02g0173800	2E-12	gi|115444527	52	*Oryza sativa*
contig0026	hypothetical protein OsJ_024219	5E-14	gi|125601160	85	*Oryza sativa*
contig0027	Os03g0571900	0.0002	gi|115453843	75	*Oryza sativa*
contig0039	hypothetical protein OsJ_028603	6E-20	gi|125606084	51	*Oryza sativa*
contig0040	hypothetical protein	5E-28	gi|147864499	98	*Vitis vinifera*
contig0041	hypothetical protein OsJ_028941	1E-08	gi|125606422	86	*Oryza sativa*
contig0046	hypothetical protein CIMG_05199	4.2	gi|119190303	38	*Coccidioides immitis*
contig0049	Os05g0529300	7.5	gi|115465043	39	*Oryza sativa*
contig0051	hypothetical protein	0.0002	gi|147768164	36	*Vitis vinifera*
contig0053	hypothetical protein OsJ_020046	3E-32	gi|125596783	77	*Oryza sativa*
contig0054	hypothetical protein CC1G_11512	8E-23	gi|116505194	66	*Coprinopsis cinerea okayama7#130*
contig0056	hypothetical protein	2E-08	gi|147811695	75	*Vitis vinifera*
contig0058	hypothetical protein DDB_0233372	0.22	gi|66819237	42	*Dictyostelium discoideum AX4*
contig0062	hypothetical protein	4E-28	gi|147791390	69	*Vitis vinifera*
contig0063	hypothetical protein AaeL_AAEL014166	2.5	gi|157103843	44	*Aedes aegypti*
contig0065	hypothetical protein	7E-11	gi|147860721	74	*Vitis vinifera*
contig0067	Os03g0772800	1E-32	gi|115455623	84	*Oryza sativa*
contig0069	hypothetical protein SS1G_11392	3.3	gi|156040828	23	*Sclerotinia sclerotiorum 1980*
contig0071	hypothetical protein OsJ_024320	2E-17	gi|125601261	68	*Oryza sativa*
contig0073	hypothetical protein	2E-28	gi|147803053	95	*Vitis vinifera*
contig0075	hypothetical protein	1E-16	gi|147853236	62	*Vitis vinifera*
contig0077	Os02g0210000	0.17	gi|115444973	50	*Oryza sativa*
contig0080	hypothetical protein PGUG_05222	4.2	gi|146412462	40	*Pichia guilliermondii ATCC 6260*
contig0082	hypothetical protein	6E-13	gi|147855522	68	*Vitis vinifera*
contig0084	hypothetical protein	1E-37	gi|147846159	89	*Vitis vinifera*
contig0085	hypothetical protein	5E-20	gi|147777804	89	*Vitis vinifera*

© University of Pretoria

contig0090	hypothetical protein	4E-38	gi|147853527	97	*Vitis vinifera*
contig0093	hypothetical protein	5E-33	gi|147768859	80	*Vitis vinifera*
contig0094	hypothetical protein PGUG_01494	7.3	gi|146419725	48	*Pichia guilliermondii ATCC 6260*
contig0099	hypothetical protein OsJ_024594	1E-21	gi|125601535	55	*Oryza sativa*
contig0100	hypothetical protein	9E-30	gi|147771668	70	*Vitis vinifera*
contig0102	hypothetical protein GLP_139_14208_19964	9.5	gi|71069367	41	*Giardia lamblia ATCC 50803*
contig0105	hypothetical protein	1E-17	gi|147855295	55	*Vitis vinifera*
contig0107	hypothetical protein	5E-30	gi|147826938	80	*Vitis vinifera*
contig0108	hypothetical protein	4E-09	gi|147820248	56	*Vitis vinifera*
contig0112	hypothetical protein	7E-09	gi|147834373	76	*Vitis vinifera*
contig0114	hypothetical protein	1E-28	gi|147771926	71	*Vitis vinifera*
contig0116	hypothetical protein	2E-20	gi|147861284	52	*Vitis vinifera*
contig0117	hypothetical protein	0.0003	gi|147861248	47	*Vitis vinifera*
contig0118	hypothetical protein	4E-48	gi|147774828	79	*Vitis vinifera*
contig0119	hypothetical protein	3E-35	gi|147780003	96	*Vitis vinifera*
contig0121	Os07g0139600	0.000001	gi|115470525	96	*Oryza sativa*
contig0124	hypothetical protein	2E-17	gi|147843854	68	*Vitis vinifera*
contig0125	hypothetical protein	2E-29	gi|6469125	66	*Cicer arietinum*
contig0127	hypothetical protein	2E-23	gi|147765903	79	*Vitis vinifera*
contig0129	predicted protein	4.4	gi|115400575	37	*Aspergillus terreus NIH2624*
contig0132	hypothetical protein CaO19_1705	7.5	gi|68468433	40	*Candida albicans SC5314*
contig0133	hypothetical protein	1E-11	gi|147841888	59	*Vitis vinifera*
Genes out of plant kingdom					
contig0005	PREDICTED: similar to B0024.8	7.2	gi|73963555	25	*Canis familiaris*
contig0007	CHK1 checkpoint homolog	6E-09	gi|89271365	85	*Xenopus tropicalis*
contig0008	GA10528-PA	2.5	gi|125984690	35	*Drosophila pseudoobscura*
contig0011	transport protein	1.9	gi|82706023	39	*Plasmodium yoelii yoelii* str. 17XNL

© University of Pretoria

contig0012	pH adaptation potassium efflux system	0.88	gi|49476293	30	*Bartonella henselae* str. Houston-1
contig0017	hypothetical protein	1E-10	gi|84579145	69	*Macaca fascicularis*
contig0019	hypothetical protein GSPATT00039354001	0.87	gi|145522917	35	*Paramecium tetraurelia*
contig0021	alpha-amylase precursor	0.67	gi|27413645	81	*Musa acuminata*
contig0031	solute carrier family 22 (organic cation transporter), member 1, isoform CRA_d	3.2	gi|119567991	35	*Homo sapiens*
contig0033	intracellular membrane-associated calcium-independent phospholipase A2 gamma isoform 1	0.88	gi|73981701	27	*Canis familiaris*
contig0043	putative role in outermembrane permeability	5.7	gi|15612203	41	*Helicobacter pylori J99*
contig0044	hypothetical protein	9.5	gi|115681474	50	*Strongylocentrotus purpuratus*
contig0045	acyl-protein synthetase, LuxE	9.4	gi|88604380	41	*Methanospirillum hungatei JF-1*
contig0048	similar to synaptic vesicle protein	7.5	gi|156555905	40	*Nasonia vitripennis*
contig0050	hypothetical protein Teth39DRAFT_1060	3E-27	gi|76795330	63	*Thermoanaerobacter ethanolicus ATCC 33223*
contig0057	diacylglycerol kinase, putative	0.85	gi|156097622	44	*Plasmodium vivax*
contig0061	putative xylose operon regulatory protein	1.1	gi|9309330	35	*Actinobacillus Actinomycetem-comitans*
contig0066	Neurohypophysial hormones, N-terminal Domain containing protein	2.5	gi|146161108	36	*Tetrahymena thermophila SB210*
contig0068	delta 12 acyl-lipid desaturase	7.2	gi|41581195	27	*Nostoc sp. 36*
contig0074	cation channel family protein	3.3	gi|118382413	42	*Tetrahymena thermophila SB210*
contig0078	putative transketolase	3.3	gi|14486703	37	*Streptococcus dysgalactiae subsp. equisimilis*
contig0079	hypothetical protein CIMG_04178	9E-08	gi|119188261	62	*Coccidioides immitis*
contig0097	probable integral membrane protein, DUF81 family	2.5	gi|71065697	46	*Psychrobacter arcticus 273-4*
contig0101	hypothetical protein GSPATT00039831001	1.9	gi|145539304	33	*Paramecium tetraurelia*

© University of Pretoria

contig0128	NADH dehydrogenase subunit 1	4.3	gi|31455604	33	*Taenia asiatica*
contig0130	probable periplasmic protein Cj0093	4.3	gi|57240715	36	*Campylobacter lari*
contig0134	broad CG11491-PB, isoform B	0.007	gi|24639132	83	*Drosophila melanogaster*
contig0135	deformed epidermal autoregulatory factor 1 homolog	5.6	gi|118344394	34	*Ciona intestinalis*
Unknown proteins					
contig0016	unnamed protein product	6E-35	gi|50312723	74	*Kluyveromyces lactis*
contig0087	unnamed protein product	2.5	gi|47216557	40	*Tetraodon nigroviridis*
contig0103	unknown protein	2E-22	gi|21539537	71	*Arabidopsis thaliana*
contig0110	protein of unknown function UPF0182	1.5	gi|134299651	41	*Desulfotomaculum reducens MI-1*
Unannotated contigs					
contig0006					
contig0009					
contig0022					
contig0028					
contig0029					
contig0032					
contig0035					
contig0042					
contig0052					
contig0055					
contig0059					
contig0095					
contig0122					
contig0123					
contig0126					

© University of Pretoria

Infected library 2

Positive gene identities

contig00016	galactinol synthase 1	1E-19	gi|146230136	93	*Verbascum phoeniceum*
contig00021	ccb206	3E-18	gi|77743872	91	*Brassica oleracea*
contig00027	auxin-repressed protein-like protein ARP1	9E-30	gi|62526583	72	*Manihot esculenta*
contig00038	Early nodulin 75 precursor (N-75) (NGM-75) unnamed protein product	8E-34	gi|128421	37	*Glycine max*
contig00053	cysteine proteinase precursor	4E-17	gi|124484383	65	*Ipomoea nil*
contig00054	metallothionein-like protein class II	4E-39	gi|125620190	82	*Nelumbo nucifera*
contig00057	dormancy/auxin associated family protein	6E-15	gi|145325447	56	*Arabidopsis thaliana*
contig00060	tonoplast intrinsic protein	6E-39	gi|115383191	93	*Triticum aestivum*
contig00062	Glycosyltransferase sugar-binding region containing DXD motif; Alpha 1,4-glycosyltransferase conserved region	7.3	gi|124359565	52	*Medicago truncatula*
contig00072	maturase	7.3	gi|6513762	30	*Orontium aquaticum*
contig00074	putative protein	0.045	gi|7630052	68	*Arabidopsis thaliana*
contig00077	Os06g0152200	2E-10	gi|115466470	52	*Oryza sativa*
contig00081	putative aquaporin PIP2-1	5E-76	gi|8071620	92	*Vitis berlandieri x Vitis rupestris*
contig00088	Elongation factor 1-alpha (EF-1-alpha) elongation factor 1-alpha	3E-38	gi|6015058	97	*Manihot esculenta*
contig00089	Os08g0436000	3E-16	gi|115476578	91	*Oryza sativa*
contig00094	extensin, putative	2E-16	gi|89257636	64	*Brassica oleracea*
contig00100	glyceraldehyde 3-phosphate dehydrogenase	6E-26	gi|51703306	93	*Daucus carota*
contig00105	late embryogenesis-like protein	2E-34	gi|3264769	84	*Prunus armeniaca*
contig00115	60S ribosomal protein L37a	2E-32	gi|58578274	95	*Capsicum chinense*
contig00117	protein kinase CK2 regulatory subunit CK2B2	7E-56	gi|11527002	87	*Zea mays*
contig00118	salt-induced AAA-Type ATPase	1E-19	gi|37894600	71	*Mesembryanthemum crystallinum*
contig00122	histidine kinase 1	7E-33	gi|116109353	86	*Medicago sativa*
contig00127	protein kinase-like protein	1E-36	gi|9542860	93	*Arabidopsis thaliana*
contig00130	type 2 diacylglycerol acyltransferase	0.005	gi|86279636	52	*Vernicia fordii*

© University of Pretoria

contig00135	glycoamidase	3E-16	gi|68137527	54	*Candida albicans*
contig00137	60S ribosomal protein L21	6E-32	gi|3885884	91	*Oryza sativa*
contig00142	wts2L	6E-10	gi|6683479	62	*Citrullus lanatus*
contig00144	molecular chaperone	8E-68	gi|111143344	87	*Ricinus communis*
contig00145	NAD-malate dehydrogenase	0.00002	gi|5123836	53	*Nicotiana tabacum*
contig00155	putative single-stranded nucleic acid binding R3H	3E-08	gi|149941242	82	*Vigna unguiculata*
contig00158	cellulose synthase-like protein D4	2E-28	gi|27372782	71	*Populus tremuloides*
contig00160	early nodulin	0.007	gi|2773251	47	*Maackia amurensis*
contig00162	P-type ATPase	4E-33	gi|14275754	88	*Hordeum vulgare*
contig00167	PDR-type ABC transporter 2	7E-54	gi|41052474	76	*Nicotiana tabacum*
contig00168	glutathione S-transferase	4E-39	gi|119633090	67	*Vitis vinifera*
contig00170	thylakoid lumen 15.0 kDa protein	0.009	gi|18423453	75	*Arabidopsis thaliana*
contig00172	fructose-bisphosphate aldolase	1E-34	gi|10800924	98	*Persea americana*
contig00176	translationally controlled tumor protein	2E-10	gi|84314025	94	*Hevea brasiliensis*
contig00179	cryptochrome 1	4E-12	gi|78217441	56	*Nicotiana sylvestris*
contig00180	S-adenosylmethionine synthetase-1	5E-33	gi|37051115	97	*Pisum sativum*
contig00185	ADP-glucose pyrophosphorylase small subunit	3E-10	gi|111660950	94	*Citrus sinensis*
contig00189	putative AAA ATPase	4E-08	gi|142942519	60	*Solanum tuberosum*
Defence-related genes					
contig00007	translationally controlled tumor protein like protein	0.0000002	gi|125987960	96	*Nicotiana tabacum*
contig00011	thaumatin	2E-20	gi|5881239	82	*Vitis riparia*
contig00043	cinnamate-4-hydroxylase	3E-16	gi|9965897	91	*Gossypium arboreum*
contig00064	metallothionein-like protein type 2	7E-41	gi|7242691	94	*Persea americana*
contig00065	AP2 domain containing protein	2E-41	gi|3264767	51	*Prunus armeniaca*
contig00073	oxysterol-binding protein	3E-28	gi|38639323	72	*Solanum tuberosum*
contig00083	thaumatin-like protein, putative	3E-29	gi|15241688	70	*Arabidopsis thaliana*
contig00095	AP2 domain containing protein	7E-27	gi|3264767	72	*Prunus armeniaca*
contig00108	profilin-like protein	5E-17	gi|84682949	100	*Cinnamomum camphora*
contig00151	putative ascorbate peroxidase	0.1	gi|110180313	89	*Cunninghamia*

© University of Pretoria

					lanceolata
contig00163	Translationally-controlled tumor protein homolog (TCTP) translationally controlled tumor protein	1E-29	gi|9979196	92	*Hevea brasiliensis*
contig00169	cysteine proteinase	8E-39	gi|148927396	85	*Elaeis guineensis*
contig00175	putative universal stress protein	2E-40	gi|45720184	89	*Cicer arietinum*
contig00187	cytochrome P450 like_TBP	2E-60	gi|1545805	88	*Nicotiana tabacum*
Hypothetical proteins					
contig00008	hypothetical protein OsI_023632	0.0000001	gi|125556794	75	*Oryza sativa*
contig00012	hypothetical protein	7E-09	gi|147843683	58	*Vitis vinifera*
contig00017	conserved hypothetical protein	3E-44	gi|90419459	77	*Aurantimonas sp. SI85-9A1*
contig00024	hypothetical protein	4E-34	gi|147791048	70	*Vitis vinifera*
contig00025	hypothetical protein	1E-37	gi|147826550	96	*Vitis vinifera*
contig00028	hypothetical protein	0.0005	gi|124484403	62	*Ipomoea nil*
contig00036	hypothetical protein	0.0006	gi|147826787	100	*Vitis vinifera*
contig00037	hypothetical protein	3E-19	gi|147799024	58	*Vitis vinifera*
contig00048	hypothetical protein	3E-16	gi|147785635	56	*Vitis vinifera*
contig00052	hypothetical protein	5E-11	gi|147843670	91	*Vitis vinifera*
contig00059	hypothetical protein OsI_036265	2.7	gi|125535818	85	*Oryza sativa*
contig00066	hypothetical protein	1E-13	gi|147794785	63	*Vitis vinifera*
contig00068	hypothetical protein	0.95	gi|147773093	46	*Vitis vinifera*
contig00071	hypothetical protein	4E-33	gi|147801246	81	*Vitis vinifera*
contig00076	hypothetical protein	8E-47	gi|147864499	94	*Vitis vinifera*
contig00082	hypothetical protein	0.0004	gi|147780480	52	*Vitis vinifera*
contig00084	Os07g0590000	7.5	gi|115473109	34	*Oryza sativa*
contig00085	hypothetical protein	3E-23	gi|147861284	59	*Vitis vinifera*
contig00086	hypothetical protein	0.077	gi|147776000	61	*Vitis vinifera*
contig00087	hypothetical protein	2E-25	gi|147780003	92	*Vitis vinifera*
contig00090	hypothetical protein	5E-30	gi|147792793	71	*Vitis vinifera*
contig00092	hypothetical protein	0.000005	gi|147781172	69	*Vitis vinifera*
contig00093	hypothetical protein	2E-22	gi|147809530	69	*Vitis vinifera*
contig00098	hypothetical protein	5E-38	gi|147785125	90	*Vitis vinifera*
contig00101	hypothetical protein	5E-39	gi|147818489	90	*Vitis vinifera*

© University of Pretoria

contig00103	hypothetical protein	0.00006	gi|147788048	88	*Vitis vinifera*
contig00109	hypothetical protein	3E-27	gi|147860721	65	*Vitis vinifera*
contig00112	hypothetical protein OsI_005523	2E-41	gi|125537895	91	*Oryza sativa*
contig00113	hypothetical protein	5E-23	gi|147801801	67	*Vitis vinifera*
contig00114	hypothetical protein	1E-48	gi|147791271	93	*Vitis vinifera*
contig00116	hypothetical protein	9E-11	gi|147854555	41	*Vitis vinifera*
contig00123	hypothetical protein	3E-16	gi|147805309	63	*Vitis vinifera*
contig00126	hypothetical protein OsJ_008424	9.5	gi|125584010	73	*Oryza sativa*
contig00128	hypothetical protein OsJ_025258	1E-14	gi|125602450	85	*Oryza sativa*
contig00131	hypothetical protein OsJ_026408	7.4	gi|125603600	44	*Oryza sativa*
contig00132	hypothetical protein	1E-17	gi|147841371	72	*Vitis vinifera*
contig00136	hypothetical protein	1E-33	gi|147854712	77	*Vitis vinifera*
contig00138	hypothetical protein	1E-09	gi|147820925	56	*Vitis vinifera*
contig00139	hypothetical protein	0.000004	gi|147819070	71	*Vitis vinifera*
contig00141	hypothetical protein	4E-27	gi|147844643	76	*Vitis vinifera*
contig00146	hypothetical protein	8E-26	gi|147771560	69	*Vitis vinifera*
contig00149	hypothetical protein OsJ_034036	5E-19	gi|125578681	55	*Oryza sativa*
contig00152	hypothetical protein OsI_020084	1E-14	gi|125553142	43	*Oryza sativa*
contig00153	hypothetical protein	2E-25	gi|147853956	100	*Vitis vinifera*
contig00159	hypothetical protein	2E-34	gi|147782293	92	*Vitis vinifera*
contig00164	hypothetical protein OsJ_003301	3E-25	gi|125571961	71	*Oryza sativa*
contig00165	hypothetical protein	5E-28	gi|147774828	84	*Vitis vinifera*
contig00171	Os04g0649600	0.0001	gi|115460968	42	*Oryza sativa*
contig00173	hypothetical protein	9E-22	gi|147785069	94	*Vitis vinifera*
contig00177	hypothetical protein	1E-24	gi|147837146	93	*Vitis vinifera*
contig00178	hypothetical protein	3E-11	gi|147801762	65	*Vitis vinifera*
contig00181	hypothetical protein	6E-29	gi|90657540	87	*Cleome spinosa*
contig00183	hypothetical protein OsI_019196	0.000001	gi|125552254	72	*Oryza sativa*

© University of Pretoria

Genes out of plant kingdom

contig00001	hypothetical protein	0.0004	gi|118116494	30	*Gallus gallus*
contig00003	hypothetical protein	7.3	gi|85075596	42	*Neurospora crassa OR74A*
contig00004	acyltransferase 3	0.061	gi|118030044	29	*Burkholderia phymatum STM815*
contig00009	hypothetical protein P700755_12072	9.5	gi|91215030	50	*Psychroflexus torquis ATCC 700755*
contig00010	Zinc finger, ZZ type family protein	0.007	gi|118380071	35	*Tetrahymena thermophila SB210*
contig00013	hypothetical protein DDBDRAFT_0216840	4.3	gi|66824867	35	*Dictyostelium discoideum AX4*
contig00015	hypothetical protein	7.3	gi|68065574	29	*Plasmodium berghei strain ANKA*
contig00018	hypothetical protein MextDRAFT_3436	1.1	gi|153898183	50	*Methylobacterium extorquens PA1*
contig00019	Serpentine Receptor, class D (delta) family member (srd-46)	0.17	gi|17567085	50	*Caenorhabditis elegans*
contig00022	hypothetical protein A55_B0061	5E-08	gi|153217596	70	*Vibrio cholerae 1587*
contig00023	hypothetical protein HNE_1677	0.17	gi|114797380	37	*Hyphomonas neptunium ATCC 15444*
contig00029	hypothetical protein TVAG_378140	4.2	gi|154420613	30	*Trichomonas vaginalis G3*
contig00030	ribosomal RNA assembly protein mis3	0.22	gi|115384518	32	*Aspergillus terreus NIH2624*
contig00032	similar to 54 kDa vacuolar H(+)-ATPase subunit	7.3	gi|118086980	37	*Gallus gallus*
contig00033	similar to M-phase phosphoprotein, mpp8	1.9	gi|125825792	35	*Danio rerio*
contig00034	hypothetical protein	3.3	gi|145552617	35	*Paramecium tetraurelia*
contig00035	predicted protein	0.035	gi|145616078	34	*Magnaporthe grisea 70-15*
contig00039	HCaRG protein	9.7	gi|118348338	36	*Tetrahymena thermophila SB210*
contig00040	heat shock protein	1.9	gi|725333	33	*Trypanosoma cruzi*
contig00041	hypothetical protein LNTAR_03759	0.51	gi|149200343	38	*Lentisphaera araneosa HTCC2155*
contig00042	C35E7.6	5.7	gi|17505875	36	*Caenorhabditis elegans*

© University of Pretoria

contig00044	CHK1 checkpoint homolog (S. pombe)	8E-09	gi|89271365	85	*Xenopus tropicalis*
contig00045	hypothetical protein GSPATT00038829001	5.6	gi|145508647	37	*Paramecium tetraurelia*
contig00046	hypothetical protein WGLp294	7.4	gi|32491043	24	*Wigglesworthia glossinidia endosymbiont of Glossina brevipalpis*
contig00049	cytochrome oxidase subunit III	0.86	gi|76359339	34	*Octopus australis*
contig00050	hypothetical protein PFF0410w	0.88	gi|86170731	28	*Plasmodium falciparum 3D7*
contig00051	hypothetical protein CHGG_02834	2	gi|116207082	32	*Chaetomium globosum CBS 148.51*
contig00056	hypothetical protein BACOVA_02878	7.2	gi|156109923	35	*Bacteroides ovatus ATCC 8483*
contig00061	hypothetical protein BU087	2.5	gi|15616709	40	*Buchnera aphidicola str. APS (Acyrthosiphon pisum)*
contig00063	predicted protein	1.5	gi|156406446	39	*Nematostella vectensis*
contig00069	hypothetical protein DDBDRAFT_0188128	9.8	gi|66805639	36	*Dictyostelium discoideum AX4*
contig00070	hypothetical protein TTHERM_01295340	3.2	gi|118375060	35	*Tetrahymena thermophila SB210*
contig00075	conserved hypothetical protein	0.002	gi|119482560	36	*Neosartorya fischeri NRRL 181*
contig00078	hypothetical protein FG00275.1	2.5	gi|46105294	37	*Gibberella zeae PH-1*
contig00079	similar to putative transmembrane transporter FLIPT 1	5.5	gi|118083486	41	*Gallus gallus*
contig00091	hypothetical protein BVU_0142	7.2	gi|150002748	45	*Bacteroides vulgates ATCC 8482*
contig00106	DNA ligase IV	7.3	gi|62859221	41	*Xenopus tropicalis*
contig00107	kinesin, putative	1.9	gi|146084573	34	*Leishmania infantum JPCM5*
contig00121	oligopeptide transporter	9.6	gi|108881208	35	*Aedes aegypti*
contig00124	cytochrome d ubiquinol oxidase, subunit II	7.3	gi|154707432	42	*Coxiella burnetii Dugway 7E9-12*
contig00125	hCG1793893	3E-08	gi|119627579	52	*Homo sapiens*
contig00129	hypothetical protein	7.3	gi|145511722	28	*Paramecium tetraurelia*
contig00133	hypothetical protein	7.3	gi|149705441	38	*Equus caballus*

© University of Pretoria

contig00140	hypothetical protein TTHERM_00566720	9.6	gi|118377791	26	*Tetrahymena thermophila SB210*
contig00147	hypothetical protein	7.3	gi|114623572	28	*Pan troglodytes*
contig00156	hypothetical protein CMM_2667	7.3	gi|148273851	30	*Clavibacter michiganensis subsp. michiganensis NCPPB*
contig00157	similar to solute carrier family 38, member 3	3.3	gi|115655641	42	*Strongylocentrotus purpuratus*
contig00161	IgG immunoreactive antigen	0.004	gi|2801529	47	*Strongyloides stercoralis*
contig00166	hypothetical protein MchlDRAFT_1971	0.4	gi|156451366	32	*Methylobacterium chloromethanicum CM4*
contig00184	hypothetical protein RRC34	4.3	gi|147919122	45	Uncultured *Methanogenic archaeon RC-I*
Unknown proteins					
contig00002	unknown protein	0.00004	gi|15221485	72	*Arabidopsis thaliana*
contig00005	unnamed protein product	0.1	gi|11994179	64	*Arabidopsis thaliana*
contig00020	expressed in cucumber hypocotyls	1E-09	gi|5139695	45	*Cucumis sativus*
contig00099	unknown protein	3E-13	gi|18405239	43	*Arabidopsis thaliana*
contig00174	At4g32960	7E-25	gi|38564256	76	*Arabidopsis thaliana*
Unannotated contigs					
contig00006					
contig00014					
contig00026					
contig00031					
contig00047					
contig00055					
contig00058					
contig00067					
contig00080					
contig00096					

© University of Pretoria

contig00097
contig00102
contig00104
contig00110
contig00111
contig00120
contig00134
contig00143
contig00148
contig00150
contig00154
contig00182
contig00186
contig00188

© University of Pretoria

Table 4. Uninfected library contigs of avocado roots infected with *Phytophthora cinnamomi.* Contigs are grouped into functional classes based on KOG annotation. Defence-related genes are indicated in functional groups by bold and italicized font.

Contig	Identity	Sequence ID
Energy production and conversion		
contig00042	hypothetical protein	gi|147792026
contig00005	conserved hypothetical protein	gi|90419459
contig00010	hypothetical protein	gi|147812476
contig00018	NADH:cytochrome b5 reductase	gi|55979111
contig00043	putative extensin	gi|148807133
contig00009	hypothetical protein	gi|147858131
Function unknown		
contig00013	Structure Of Farnesyl Pyrophosphate Synthetase Chain	gi|1942048
contig00030	hypothetical protein OsI_013039	gi|125545667
General function prediction only		
contig00014	Chain A, Solution Structrue Of The Zf-An1 Domain From Arabidopsis Thaliana At2g36320 Protein	gi|159163297
contig00039	putative transcription factor	gi|14582465
Inorganic ion transport and metabolism		
contig00032	hypothetical protein	gi|147775208
contig00033	***metallothionein-like protein type 2***	gi|7242691
Intracellular trafficking, secretion, and vesicular transport		
contig00008	secreted antigen 1	gi|118442730
Lipid transport and metabolism		
contig00036	hypothetical protein	gi|147834261
contig00004	hypothetical protein MGG_06317	gi|145608858

© University of Pretoria

Posttranslational modification, protein turnover, chaperones		
contig00027	LRX1 (LEUCINE-RICH REPEAT/EXTENSIN 1); protein binding / structural constituent of cell wall	gi|15221178
contig00038	oil palm polygalacturonase allergen PEST472	gi|34223513
contig00015	hypothetical protein	gi|147822174
RNA processing and modification		
contig00020	hypothetical protein COSY_0553	gi|148244700
contig00037	RNA and export factor-binding protein, putative	gi|145334661
contig00001	Magnolia denudata 26S ribosomal RNA gene, partial sequence	gi|1545805
contig00028	ZIP family metal transporter	gi|82581289
Secondary metabolites biosynthesis, transport and catabolism		
contig00016	Na(+)-translocating NADH-quinone reductase subunit B	gi|149177061
contig00023	hypothetical protein PARMER_02310	gi|154492673
Signal transduction mechanisms		
contig00011	Camellia sinensis 18S Ribosomal RNA (from rRNA database)	gi|147826550
contig00025	hypothetical protein	gi|147815590
contig00026	polyprotein	gi|156447514
contig00031	protein kinase family protein	gi|30696480
contig00024	CHR4/MI-2-LIKE (chromatin remodeling 4); ATP binding / DNA binding / chromatin binding / helicase	gi|145358870
Transcription		
contig00006	Late embryogenesis abundant protein Lea5 Lea5 protein	gi|22653808
contig00029	hypothetical protein	gi|147781769
contig00034	hypothetical protein	gi|147841226
contig00035	hypothetical protein	gi|147769159

© University of Pretoria

Translation, ribosomal structure and biogenesis		
contig00002	T6D22.2	gi|8778823
contig00003	Camellia sinensis 18S Ribosomal RNA (from rRNA database)	gi|50312723
contig00017	60S ribosomal protein L36 homolog	gi|19979631
contig00021	hypothetical protein	gi|147798057

© University of Pretoria

Table 5. Infected library 1 contigs of avocado roots infected with *Phytophthora cinnamomi*. Contigs are grouped into functional classes based on KOG annotation. Defence-related genes are indicated in functional groups by bold and italicized font.

Contig	Identity	Sequence ID
Amino acid transport and metabolism		
contig0047	bZIP transcription factor bZIP80	gi|113367194
contig0077	Os02g0210000	gi|115444973
contig0007	CHK1 checkpoint homolog (S. pombe)	gi|89271365
contig0011	transport protein	gi|82706023
Carbohydrate transport and metabolism		
contig0041	hypothetical protein OsJ_028941	gi|125606422
contig0100	hypothetical protein	gi|147771668
contig0078	putative transketolase	gi|14486703
contig0112	hypothetical protein	gi|147834373
Cell cycle control, cell division, chromosome partitioning		
contig0083	putative DnaJ protein	gi|145688390
contig0102	hypothetical protein GLP_139_14208_19964	gi|71069367
contig0023	Os02g0173800	gi|115444527
Cell wall/membrane/envelope biogenesis		
contig0084	hypothetical protein	gi|147846159
contig0104	sugar binding / transferase, transferring glycosyl groups	gi|145335796
Chromatin structure and dynamics		
contig0117	hypothetical protein	gi|147861248
contig0132	hypothetical protein CaO19_1705	gi|68468433
contig0039	hypothetical protein OsJ_028603	gi|125606084
Cytoskeleton		
contig0113	ACT1	gi|149938964
contig0139	extensin-like protein	gi|791148

© University of Pretoria

contig0020	***leucine-rich repeat resistance protein-like protein***	gi|14626935
contig0093	hypothetical protein	gi|147768859
contig0008	GA10528-PA	gi|125984690
contig0066	Neurohypophysial hormones, N-terminal Domain containing protein	gi|146161108
contig0056	hypothetical protein	gi|147811695
contig0045	acyl-protein synthetase, LuxE	gi|88604380
Energy production and conversion		
contig0092	ATP synthase epsilon chain, mitochondrial	gi|15217996
contig0120	Auxin-induced protein PCNT115 auxin-induced protein	gi|728744
contig0049	Os05g0529300	gi|115465043
contig0058	hypothetical protein DDB_0233372	gi|66819237
contig0067	Os03g0772800	gi|115455623
Extracellular structures		
contig0109	***drought-induced protein***	gi|14994231
Function unknown		
contig0024	Elongation factor 2 (EF-2) elongation factor 2	gi|6015065
contig0030	Major latex allergen Hev b 5 latex allergen pI 3.5 latex allergen	gi|7387766
contig0001	hypothetical protein	gi|147777692
contig0004	hypothetical protein PFE0570w	gi|124506149
contig0014	hypothetical protein	gi|147865432
contig0053	hypothetical protein OsJ_020046	gi|125596783
contig0094	hypothetical protein PGUG_01494	gi|146419725
contig0005	PREDICTED: similar to B0024.8	gi|73963555
contig0135	deformed epidermal autoregulatory factor 1 homolog	gi|118344394

© University of Pretoria

General function prediction only		
contig0018	dormancy-associated protein/auxin-repressed protein	gi|148872940
contig0060	pantothenate kinase family protein	gi|18418688
contig0002	hypothetical protein FG01701.1	gi|46109638
contig0027	Os03g0571900	gi|115453843
contig0116	hypothetical protein	gi|147861284
contig0127	hypothetical protein	gi|147765903
contig0031	solute carrier family 22 (organic cation transporter), member 1, isoform CRA_d	gi|119567991
contig0043	putative role in outermembrane permeability	gi|15612203
contig0057	diacylglycerol kinase, putative	gi|156097622
contig0097	probable integral membrane protein, DUF81 family	gi|71065697
contig0128	NADH dehydrogenase subunit 1	gi|31455604
contig0134	broad CG11491-PB, isoform B	gi|24639132
contig0129	predicted protein	gi|115400575
Inorganic ion transport and metabolism		
contig0076	metallothionein-like protein type 2	gi|7242691
contig0063	hypothetical protein AaeL_AAEL014166	gi|157103843
Intracellular trafficking, secretion, and vesicular transport		
contig0081	putative AP3-complex beta-3A adaptin subunit	gi|16604671
contig0131	protein binding / zinc ion binding	gi|145339866
contig0054	hypothetical protein CC1G_11512	gi|116505194
contig0012	pH adaptation potassium efflux system e	gi|49476293
contig0021	alpha-amylase precursor	gi|27413645
Lipid transport and metabolism		
contig0010	hypothetical protein DDBDRAFT_0202116	gi|66827755
contig0101	hypothetical protein GSPATT00039831001	gi|145539304

© University of Pretoria

Nucleotide transport and metabolism		
contig0061	putative xylose operon regulatory protein	gi|9309330
contig0019	hypothetical protein GSPATT00039354001	gi|145522917
Posttranslational modification, protein turnover, chaperones		
contig0111	70 kDa peptidyl-prolyl isomerase (Peptidyl-prolyl cis-trans isomerase) (PPIase) (Rotamase) peptidylprolyl isomerase	gi|3023751
contig0073	hypothetical protein	gi|147803053
contig0105	hypothetical protein	gi|147855295
contig0079	hypothetical protein CIMG_04178	gi|119188261
contig0048	PREDICTED: similar to synaptic vesicle protein	gi|156555905
contig0037	pentatricopeptide (PPR) repeat-containing protein	gi|15218705
contig0013	hypothetical protein	gi|147800856
contig0074	cation channel family protein	gi|118382413
RNA processing and modification		
contig0072	Putative small nuclear ribonucleoprotein G	gi|14192871
contig0125	hypothetical protein	gi|6469125
contig0068	delta 12 acyl-lipid desaturase	gi|41581195
contig0082	hypothetical protein	gi|147855522
contig0017	hypothetical protein	gi|84579145
contig0025	ENOD2f	gi|10198184
contig0124	hypothetical protein	gi|147843854
Secondary metabolites biosynthesis, transport and catabolism		
contig0038	caffeoyl-CoA 3-O-methyltransferase, putative	gi|18416703
contig0065	hypothetical protein	gi|147860721
contig0118	hypothetical protein	gi|147774828
contig0099	hypothetical protein OsJ_024594	gi|125601535

© University of Pretoria

Signal transduction mechanisms		
contig0036	putative polyprotein	gi|54291863
contig0088	***seven transmembrane protein Mlo6***	gi|13784985
contig0106	***pathogenesis-related protein PsemI***	gi|11080640
contig0015	hypothetical protein	gi|147770332
contig0026	hypothetical protein OsJ_024219	gi|125601160
contig0046	hypothetical protein CIMG_05199	gi|119190303
contig0051	hypothetical protein	gi|147768164
contig0062	hypothetical protein	gi|147791390
contig0069	hypothetical protein SS1G_11392	gi|156040828
contig0080	hypothetical protein PGUG_05222	gi|146412462
contig0107	hypothetical protein	gi|147826938
contig0114	hypothetical protein	gi|147771926
contig0044	PREDICTED: hypothetical protein	gi|115681474
contig0064	putative glycine rich protein	gi|20152613
contig0089	proline-rich extensin, putative	gi|15222668
contig0096	PLA IIB/PLP6 (Patatin-like protein 6); nutrient reservoir	gi|15225054
contig0137	extensin	gi|1486263
contig0108	hypothetical protein	gi|147820248
contig0050	hypothetical protein Teth39DRAFT_1060	gi|76795330
Transcription		
contig0091	putative plastidic adenylate transporter	gi|61651614
contig0070	***4-coumarate-CoA ligase-like protein***	gi|29893225
contig0003	hypothetical protein	gi|147835464
contig0075	hypothetical protein	gi|147853236
contig0133	hypothetical protein	gi|147841888
contig0033	PREDICTED: similar to intracellular membrane-associated calcium-independent phospholipase A2 gamma isoform 1	gi|73981701
contig0071	hypothetical protein OsJ_024320	gi|125601261

© University of Pretoria

Translation, ribosomal structure and biogenesis		
contig0034	alpha-amylase inhibitor-like precursor	gi|56237708
contig0086	glycyl-tRNA synthetase	gi|110740563
contig0098	tRNA synthetase class II (G, H, P and S) family protein	gi|15228692
contig0115	60S ribosomal protein L10 (QM protein homolog) 60S ribosomal protein L10	gi|18203445
contig0136	60S ribosomal protein L10 60S ribosomal protein L10	gi|18203270
contig0040	hypothetical protein	gi|147864499
contig0085	hypothetical protein	gi|147777804
contig0090	hypothetical protein	gi|147853527
contig0119	hypothetical protein	gi|147780003
contig0121	Os07g0139600	gi|115470525
contig0130	probable periplasmic protein Cj0093	gi|57240715

© University of Pretoria

Table 6. Infected library 2 contigs of avocado roots infected with *Phytophthora cinnamomi.* Contigs are grouped into functional classes based on KOG annotation. Defence-related genes are indicated in functional groups by bold and italicized font.

Contig	Identity	Sequence ID
Amino acid transport and metabolism		
contig00074	putative protein	gi|7630052
contig00112	hypothetical protein OsI_005523	gi|125537895
contig00044	CHK1 checkpoint homolog (S. pombe)	gi|89271365
Carbohydrate transport and metabolism		
contig00016	galactinol synthase 1	gi|146230136
contig00060	tonoplast intrinsic protein	gi|115383191
contig00081	putative aquaporin PIP2-1	gi|8071620
contig00100	glyceraldehyde 3-phosphate dehydrogenase	gi|51703306
contig00172	fructose-bisphosphate aldolase	gi|10800924
contig00068	hypothetical protein	gi|147773093
contig00114	hypothetical protein	gi|147791271
contig00128	hypothetical protein OsJ_025258	gi|125602450
contig00152	hypothetical protein OsI_020084	gi|125553142
contig00159	hypothetical protein	gi|147782293
contig00157	PREDICTED: similar to solute carrier family 38, member 3	gi|115655641
Cell cycle control, cell division, chromosome partitioning		
contig00145	NAD-malate dehydrogenase	gi|5123836
contig00017	conserved hypothetical protein	gi|90419459
contig00071	hypothetical protein	gi|147801246
contig00013	hypothetical protein DDBDRAFT_0216840	gi|66824867
contig00078	hypothetical protein FG00275.1	gi|46105294
contig00176	translationally controlled tumor protein	gi|84314025
contig00007	***translationally controlled tumor protein like protein***	gi|125987960

© University of Pretoria

contig00163	***Translationally-controlled tumor protein homolog (TCTP) translationally controlled tumor protein***	gi|9979196
contig00177	hypothetical protein	gi|147837146
contig00066	hypothetical protein	gi|147794785
Cell wall/membrane/envelope biogenesis		
contig00021	ccb206	gi|77743872
contig00185	ADP-glucose pyrophosphorylase small subunit	gi|111660950
contig00019	Serpentine Receptor, class D (delta) family member (srd-46)	gi|17567085
contig00070	hypothetical protein TTHERM_01295340	gi|118375060
Coenzyme transport and metabolism		
contig00180	S-adenosylmethionine synthetase-1	gi|37051115
contig00042	C35E7.6	gi|17505875
Cytoskeleton		
contig00027	auxin-repressed protein-like protein ARP1	gi|62526583
contig00038	Early nodulin 75 precursor (N-75) (NGM-75) unnamed protein product	gi|128421
contig00057	dormancy/auxin associated family protein	gi|145325447
contig00108	***profilin-like protein***	gi|84682949
contig00010	Zinc finger, ZZ type family protein	gi|118380071
contig00030	ribosomal RNA assembly protein mis3	gi|115384518
Energy production and conversion		
contig00170	thylakoid lumen 15.0 kDa protein	gi|18423453
contig00052	hypothetical protein	gi|147843670
contig00146	hypothetical protein	gi|147771560
contig00051	hypothetical protein CHGG_02834	gi|116207082
contig00161	IgG immunoreactive antigen	gi|2801529
contig00018	hypothetical protein MextDRAFT_3436	gi|153898183
Function unknown		
contig00179	cryptochrome 1	gi|78217441
contig00090	hypothetical protein	gi|147792793

© University of Pretoria

contig00126	hypothetical protein OsJ_008424	gi|125584010
contig00132	hypothetical protein	gi|147841371
contig00139	hypothetical protein	gi|147819070
contig00015	hypothetical protein	gi|68065574
contig00022	hypothetical protein A55_B0061	gi|153217596
contig00023	hypothetical protein HNE_1677	gi|114797380
contig00032	PREDICTED: similar to 54 kDa vacuolar H(+)-ATPase subunit	gi|118086980
contig00034	hypothetical protein	gi|145552617
contig00041	hypothetical protein LNTAR_03759	gi|149200343
contig00079	PREDICTED: similar to putative transmembrane transporter FLIPT 1	gi|118083486
contig00091	hypothetical protein BVU_0142	gi|150002748
contig00106	DNA ligase IV	gi|62859221
General function prediction only		
contig00062	Glycosyltransferase sugar-binding region containing DXD motif; Alpha 1,4-glycosyltransferase conserved region	gi|124359565
contig00072	maturase	gi|6513762
contig00105	late embryogenesis-like protein	gi|3264769
contig00142	wts2L	gi|6683479
contig00151	***putative ascorbate peroxidase***	gi|110180313
contig00024	hypothetical protein	gi|147791048
contig00028	hypothetical protein	gi|124484403
contig00085	hypothetical protein	gi|147861284
contig00086	hypothetical protein	gi|147776000
contig00103	hypothetical protein	gi|147788048
contig00183	hypothetical protein OsI_019196	gi|125552254
contig00001	PREDICTED: hypothetical protein	gi|118116494
contig00035	predicted protein	gi|145616078
contig00049	cytochrome oxidase subunit III	gi|76359339

© University of Pretoria

contig00129	hypothetical protein	gi|145511722
contig00084	Os07g0590000	gi|115473109
Inorganic ion transport and metabolism		
contig00054	metallothionein-like protein class II	gi|125620190
contig00162	P-type ATPase	gi|14275754
contig00064	***metallothionein-like protein type 2***	gi|7242691
contig00059	hypothetical protein OsI_036265	gi|125535818
contig00082	hypothetical protein	gi|147780480
contig00033	PREDICTED: similar to M-phase phosphoprotein, mpp8	gi|125825792
contig00061	hypothetical protein BU087	gi|15616709
contig00107	kinesin, putative	gi|146084573
contig00009	hypothetical protein P700755_12072	gi|91215030
Intracellular trafficking, secretion, and vesicular transport		
contig00122	histidine kinase 1	gi|116109353
contig00065	***AP2 domain containing protein***	gi|3264767
contig00083	***thaumatin-like protein, putative***	gi|15241688
contig00175	***putative universal stress protein***	gi|45720184
contig00136	hypothetical protein	gi|147854712
contig00141	hypothetical protein	gi|147844643
contig00046	hypothetical protein WGLp294	gi|32491043
Lipid transport and metabolism		
contig00130	type 2 diacylglycerol acyltransferase	gi|86279636
contig00073	***oxysterol-binding protein***	gi|38639323
contig00092	hypothetical protein	gi|147781172
contig00004	acyltransferase 3	gi|118030044
contig00056	hypothetical protein BACOVA_02878	gi|156109923
contig00124	cytochrome d ubiquinol oxidase, subunit II	gi|154707432
contig00039	HCaRG protein	gi|118348338

© University of Pretoria

Posttranslational modification, protein turnover, chaperones		
contig00053	cysteine proteinase precursor	gi|124484383
contig00077	Os06g0152200	gi|115466470
contig00118	salt-induced AAA-Type ATPase	gi|37894600
contig00144	molecular chaperone	gi|111143344
contig00168	glutathione S-transferase	gi|119633090
contig00189	putative AAA ATPase	gi|142942519
contig00169	***cysteine proteinase***	gi|148927396
contig00012	hypothetical protein	gi|147843683
contig00093	hypothetical protein	gi|147809530
contig00149	hypothetical protein OsJ_034036	gi|125578681
contig00153	hypothetical protein	gi|147853956
contig00164	hypothetical protein OsJ_003301	gi|125571961
contig00178	hypothetical protein	gi|147801762
contig00050	hypothetical protein PFF0410w	gi|86170731
contig00069	hypothetical protein DDBDRAFT_0188128	gi|66805639
contig00029	hypothetical protein TVAG_378140	gi|154420613
Replication, recombination and repair		
contig00045	hypothetical protein GSPATT00038829001	gi|145508647
contig00140	hypothetical protein TTHERM_00566720	gi|118377791
RNA processing and modification		
contig00089	Os08g0436000	gi|115476578
contig00155	putative single-stranded nucleic acid binding R3H	gi|149941242
contig00011	***thaumatin***	gi|5881239
contig00008	hypothetical protein OsI_023632	gi|125556794
contig00113	hypothetical protein	gi|147801801
contig00138	hypothetical protein	gi|147820925
contig00156	hypothetical protein CMM_2667	gi|148273851

© University of Pretoria

Secondary metabolites biosynthesis, transport and catabolism		
contig00167	PDR-type ABC transporter 2	gi|41052474
contig00043	***cinnamate-4-hydroxylase***	gi|9965897
contig00101	hypothetical protein	gi|147818489
contig00109	hypothetical protein	gi|147860721
contig00165	hypothetical protein	gi|147774828
Signal transduction mechanisms		
contig00127	protein kinase-like protein	gi|9542860
contig00158	cellulose synthase-like protein D4	gi|27372782
contig00160	early nodulin	gi|2773251
contig00095	***AP2 domain containing protein***	gi|3264767
contig00187	***cytochrome P450 like_TBP***	gi|1545805
contig00025	hypothetical protein	gi|147826550
contig00037	hypothetical protein	gi|147799024
contig00048	hypothetical protein	gi|147785635
contig00131	hypothetical protein OsJ_026408	gi|125603600
contig00181	hypothetical protein	gi|90657540
contig00040	heat shock protein	gi|725333
contig00121	oligopeptide transporter	gi|108881208
contig00133	PREDICTED: hypothetical protein	gi|149705441
contig00063	predicted protein	gi|156406446
contig00117	protein kinase CK2 regulatory subunit CK2B2	gi|11527002
contig00094	extensin, putative	gi|89257636
contig00135	glycoamidase	gi|68137527
contig00166	hypothetical protein MchlDRAFT_1971	gi|156451366
contig00173	hypothetical protein	gi|147785069
Transcription		
contig00098	hypothetical protein	gi|147785125
contig00123	hypothetical protein	gi|147805309
contig00003	hypothetical protein	gi|85075596

© University of Pretoria

contig00075	conserved hypothetical protein	gi|119482560
contig00125	hCG1793893	gi|119627579
Translation, ribosomal structure and biogenesis		
contig00088	Elongation factor 1-alpha (EF-1-alpha) elongation factor 1-alpha	gi|6015058
contig00115	60S ribosomal protein L37a	gi|58578274
contig00137	60S ribosomal protein L21	gi|3885884
contig00036	hypothetical protein	gi|147826787
contig00076	hypothetical protein	gi|147864499
contig00087	hypothetical protein	gi|147780003
contig00116	hypothetical protein	gi|147854555
contig00171	Os04g0649600	gi|115460968
contig00147	PREDICTED: hypothetical protein	gi|114623572
contig00184	hypothetical protein RRC34	gi|147919122

© University of Pretoria

Table 7. Contigs of *Phytophthora cinnamomi* infected R0.09 avocado roots grouped into functional classes according to GO database. 44.5% of contigs fall into the unknown classification. 3% of the contigs link directly with a stress response.

Gene ontology	Number of contigs
Unknown	165
Other	48
ATP binding	7
Biological process - cytoplasm	20
Cellular component	39
Response to stress	12
RNA binding	5
Cell wall related	10
Structural constituent of ribosome - translation - ribosome	12
Transcription factor activity - regulation of transcription	7
Transferase activity/cell wall biogenesis	4
Translation elongation factor activity - translation factor activity, nucleic acid binding	3
Transporter activity - transport	3
Water channel activity - transport - membrane	4
Protein binding	10
Mitochondrion	9
Kinase activity	4
Protein folding – cellular component	3
Membrane	3
Lipid binding - lipid transport	2
Total	**370**

© University of Pretoria

Table 8. Percentage of contigs associated to *Persea americana* from the *Phytophthora cinnamomi* infected R0.09 avocado roots sequence data. A large number of transcripts showed homology to the grape vine in all three cDNA libraries. In total there were only 4 transcripts that showed homology to *P. americana.* The 'out of plant kingdom' group represents data that showed homology to non-relevant species such as humans, mice or other bacteria etc. The 'other' group is a compilation of various plant species

Species	Uninfected	Infected library 1	Infected library 2	total
Vitis vinifera	13	30	39	82
Arabidopsis thaliana	6	13	9	28
Persea americana	1	1	2	4
Oryza sativa	0	14	15	29
Nicotiana tabacum	0	1	4	5
Out of plant kingdom	6	28	50	84
Other	17	52	70	139
Total	**43**	**139**	**189**	**371**

© University of Pretoria

CHAPTER 3

Quantification of avocado defence-related genes in response to *Phytophthora cinnamomi*

© University of Pretoria

3.1 Introduction

Despite the importance of avocado and a 60 year attempt to unravel the host pathogen interaction, Phytophthora root rot (PRR) is still the most limiting disease in avocado production. Currently the information surrounding PRR and tolerance to *Phytophthora cinnamomi* Rands. is based on the analysis of root exudates (Botha & Kotze, 1989), chemical analysis of roots (Sánchez-Pérez *et al.*, 2009), the application of chemicals to aid in suppression of the pathogen (Bekker *et al.*, 2006), and biochemical studies (García-Pineda *et al.*, 2010). There have also been histological studies performed on roots infected with *P. cinnamomi* in order to try and understand the plant pathogen interaction (Phillips *et al.*, 1987). It was observed that necrophylactic periderm and periclinal cell wall division occurred, which limited the pathogens progress but did not affect the viability of the pathogen or reduce its ability to infect the host plant. *Phytophthora cinnamomi* infect the plants roots via motile zoospores present in the soil. The attraction of zoospores was investigated by Botha and Kotze in 1989 and it was found to be influenced by the composition of amino acids in root exudates (Botha & Kotze, 1989). A composition of 14 amino acids attracted zoospores more than sugars or any 1 amino acid alone. In another study by Sánchez-Pérez and colleagues, crude root exudates were tested for *P. cinnamomi* mycelial inhibition and subsequently the compound known as stigmastan-3, 5-diene was identified as the inhibitory compound (Sánchez-Pérez *et al.*, 2009). García-Pineda (2010) and colleagues investigated reactive oxygen species (ROS) formation and the role of nitric oxide (NO) against *P. cinnamomi* (García-Pineda *et al.*, 2010). The authors observed an increase in ROS and NO levels and deduced that the increase in ROS observed may assist in weakening host tissue early in infection with the sharp increase in NO possibly resulting in salicylic acid (SA) accumulation. The effect of SA on root colonisation was also investigated and indicated that decreased root colonisation was associated with increased SA levels. The production of NO and ROS have previously been demonstrated to activate cell death. These

© University of Pretoria

early attempts on investigating the interaction between avocado and *P. cinnamomi* have illustrated the complexity of the defence response, highlighting the need for the molecular elucidation of defence genes.

Genetic information available for avocado is limited. To date molecular research has comprised of genetic relationship studies and the molecular characterization of the fruit and flowers. There has been some gene characterization of avocado fruit ripening genes (Cass *et al.*, 1990; Chernys & Zeevaart, 2000; Christoffersen *et al.*, 1984; Hammond-Kosack & Jones, 1997; Kupke *et al.*, 2003). The greater part of molecular detail exists due to a continuous effort in marker development to assist in either elucidating genetic relationships amongst scions (Acheampong *et al.*, 2008; Chanderbali *et al.*, 2008; Chang *et al.*, 1993; Clegg & Davis, 1989; Davis *et al.*, 1998; Mhameed *et al.*, 1997), or scion improvement (Chen *et al.*, 2007; Clegg *et al.*, 1992; Clegg *et al.*, 1999; Lavi *et al.*, 1991). There is currently a preliminary genetic map available based on microsatellites, random amplified polymorphic DNA (RAPD) markers and DNA fingerprint (DFP) markers (Sharon *et al.*, 1997).

AFLP markers have been used to characterise PRR tolerance at the University of California Riverside (UCR) (Douhan *et al.*, 2011). The study concluded that resistance mechanisms vary between tolerant cultivars and there was no marker pattern that could be linked to tolerance. The most recent molecular development in the fight against PRR was the identification of 70 microsatellite markers that were developed from over 8000 ESTs in the hope of aiding in marker assisted breeding against PRR. The ESTs were however from a floral gene database generated for comparative genomics research of basal angiosperms. Their efficacy has yet to be tested for use in identifying tolerant rootstocks, but it is known that they amplify across all avocado varieties and can be used for investigation of genetic relations (Albert *et al.*, 2005; Borrone *et al.*, 2007). ESTs have previously been generated from fruiting and flowering organs of avocado

© University of Pretoria

with a lack of EST data available for avocado rootstocks, more specifically a lack of information on the avocado/ *P. cinnamomi* interaction.

The avocado/ *P. cinnamomi* interaction has not previously been elucidated on a molecular level. Current knowledge is based on research of the non-host plant, *Arabidopsis*. A study conducted on *Arabidopsis* infected with *P. cinnamomi* revealed that ROS induction, HR activation, lignin synthesis and callose production was initiated upon infection. The non-host showed activation of the ethylene and jasmonic acid pathways and only a minor involvement of the SA pathway (Rookes *et al.*, 2008) in contrast to the study conducted by García-Pineda *et al* (2010) on avocado which indicated that SA is a major inhibitor of pathogen colonisation. Macroscopic changes such as callose production have also been observed during *P. cinnamomi* infection in maize (Hinch & Clarke, 1982) and although model plants like *Arabidopsis* provide an insight to defence responses there are differences between non-host and host-specific defence responses. In order to fully understand tolerance in avocado it is important to conduct molecular level studies on the host specific interaction between *P. americana* and *P. cinnamomi*.

In order to understand the transcriptional changes that occur in the plant a reliable and sensitive technique is required to study the expression of genes in response to a pathogen. A good method that has been successful in this regard is the real-time polymerase chain reaction (Gachon *et al.*, 2004). The identification of genes and their expression may help in understanding resistance mechanisms. In this study, genes associated with resistance were identified in avocado root infection libraries and are quantified using qRT-PCR.

© University of Pretoria

3.2 Materials & Methods

3.2.1 Avocado inoculation with *P. cinnamomi* & harvesting of root material

Nine month old tolerant R0.09 clonal avocado plantlets were provided by Westfalia Technological Services (Tzaneen, South Africa) and inoculated with *P. cinnamomi* mycelia. A total of 33 g of mycelia was homogenised in 65 L of distilled water giving a final concentration of 0.5 g/L. This was then mixed into 112 kg of vermiculite in a mistbed. Plantlets were randomly grounded in vermiculite and constantly irrigated over a period of six weeks. Root material was harvested at 0 hour (uninfected), 3, 6, 12, 24 and 48 hours post infection (hpi), snap frozen in liquid nitrogen and stored on dry ice (-78 °C) until the root material could be transported back to a -80 °C freezer. A subset of plants was left in the mistbed for six weeks as a positive control for disease.

3.2.2 Generation of cDNA template

RNA isolations were done using the CTAB method (Chang *et al.*, 1993). Roots were ground in liquid nitrogen and 2-3 g of root material was used per RNA extraction. The Chloroform: isoamyl alcohol wash step was repeated 6 times followed by washing with ethanol thrice. Total RNA concentrations were quantified using the Nanodrop ND-100 Spectrophotometer (Nanodrop Technologies, Inc., Montchanin, USA) and verified on a 2% non-denaturing TAE agrose gel. Three technical replicates per biological replicate were performed for RNA isolation.

cDNA synthesis was performed for qRT-PCR with starting material of 1 - 2 µg total RNA of each time point. The ImProm-II™ single strand cDNA system from Promega (Promega Corporation, Madison, Wisconsin, USA) was used in conjunction with random hexamer primers from

© University of Pretoria

Invitrogen (Invitrogen Life Technologies, Mississauga, ON, Canada). Reverse transcription was carried out under the following conditions: 25°C for 5 min, 42°C for 60 min and 70°C for 10 min. DNA contamination was assessed in cDNA as previously described in chapter 2.

3.2.3 Quantitative gene expression analysis

Ten genes were selected for gene expression analysis and included: *Thaumatin, thaumatin-like*, *metallothionein-like, leucine-rich repeat resistance protein-like,* pathogenesis-related protein *Pseml*, *putative universal stress, profilin-like*, transmembrane protein *MLO,* the oxysterol-binding gene and *cytochrome P450-like TBP* (TATA box binding protein) (Table 1). Reference genes chosen for the study were *actin* and *18s* rRNA genes. The time points used to analyze the gene expression were 0 hours prior infection and, 3, 6, 12, 24 and 48 hours post infection.

Primers were designed using Primer Designer 4 for Windows, version 4.2©, (scientific and educational software, Durham, NC) based on sequence homology. Primers were designed to amplify products of no more than 150 bp, the selection was then auto adjusted using more aggressive criteria and primers were chosen within a 58- 61 °C range. Primers were synthesised by Southern Cross Biotechnology (Cape Town, South Africa). The thaumatin-like transcript had the highest annealing temperature of 65°C. Different annealing temperatures were analysed on separate PCR plates using an inter-run calibrator so that comparisons may be made between plates.

Quantitative PCR was carried out using the Bio-Rad CFX96 real-time PCR detection system (Bio-Rad Laboratories, Hercules, CA). Reactions were performed in a 20 µl tube containing 5 µl of each cDNA sample, 10 µl of iQ SYBR Green supermix (Bio-Rad), 1 µl of each set of fragment specific forward and reverse primers and 3 µl SABAX water.

© University of Pretoria

Thermocycling was carried out at 95°C for 10 min, followed by 55 cycles of 95°C for 10 sec, fragment specific annealing temperatures (Table 2) for 15 sec and elongation of 72°C for 15 sec. Three independent biological replicates were used with three technical replicates of each. The specificity of each primer pair was thoroughly investigated by standard PCR before the quantitative PCR was conducted and then verified by the presence of a single melting temperature peak. For the qPCR, the cycle threshold (Ct) values were automatically determined using the accompanying Bio-Rad CFX manager (Bio-Rad CFX Manager™ version 1.5).

3.2.4 Statistical analysis

A Student T-test was carried out to determine significant differences between gene expression levels for quantitative gene expression analysis. Statistical analysis was performed using the JMP® program version 9.0.0 (SAS Institute, Inc., Cary, NC) with a 95% confidence interval.

3.3 Results

3.3.1 Symptoms on plants

Inoculation of the avocado plantlets resulted in the development of PRR symptoms within six weeks. The plantlets showed signs of wilting and deterioration in overall plant health. Upon inspection of the roots it was found that vascular discolouration had begun and the young feeder roots that had been infected showed signs of necrosis. *Phytophthora cinnamomi* was re-isolated from diseased roots, proving Koch's postulates. These results indicate that the infection method was successful and caused disease.

© University of Pretoria

3.3.2 cDNA results

RNA extraction yield was very high and of good quality. Distinct 28S and 18S bands were visualised for all RNA extractions with no severe degradation being visualized (Figure 1). Successful cDNA synthesis resulted in a smear between 500 bp and 3 kb.

Amplification using the F3H primers indicated that there was no DNA contamination yielding the expected amplicon of ~240 bp for cDNA compared to the 1200 bp fragment expected for genomic DNA. The difference in band size is due to the removal of the intron sequence in cDNA. All samples were negative for the 1200 bp band but positive for the smaller cDNA band.

Quantitative gene expression analysis

Ten avocado root transcripts were quantified at six time points, 0, 3, 6, 12, 24 and 48 hours. Signal intensities were strong for all genes with Ct values below the non template control fluorescence. The genes quantified all indicated 100% efficiency with good repeatable Ct values across technical replicates. The expression analysis was anchored using two endogenous control genes (18s and actin). The expression data was then compared against the pyrosequencing data (chapter 2) which revealed that six of the 10 genes showed the highest expression at a time point belonging to the library in which the transcript was identified (Table 3).

Thaumatin expression was significantly greater at 48 hpi (1.1) as oppose to the uninfected (0.4), as well as the 3 & 6 hpi (Figure 2a). The expression pattern indicated that thaumatin was only regulated in response to *P. cinnamomi* by 12 hpi and increased by nearly threefold over a 36 hour period. Thaumatin levels were significantly higher in the later infection time points when compared to the earlier time points- thus correlating with the pyrosequencing data.

© University of Pretoria

The pathogenesis-related (PR-10) *pseml* gene showed significant increases in expression at 6 & 24 hpi. At 24 hpi *pseml* reached the highest expression level of 1.5 when compared to all time points (Figure 2b). By 48 hpi *pseml* expression had decreased significantly to 0.1, reaching levels comparable to 3 hpi.

Cytochrome P450-like TBP (TATA box binding protein) showed a significant early response at 3 hr after infection with *P. cinnamomi* reacting with an increase of ten-fold. At 6 hpi the gene was significantly down-regulated followed by a substantial increase at 12 hp, a similar level found at 3 hpi. The gene was then significantly down-regulated to 1.0 at 24 hpi and remained unchanged at 48 hpi (Figure 2c). Cytochrome P450-like TBP levels were constantly up- and down-regulated showing significant variation over time points. The data was consistent with the pyrosequencing data for this transcript.

The gene encoding for a metallothionein-like protein was constitutively expressed at 0.6 and showed no significant changes in expression over the first 6 hpi. This was however followed by a significant increase in the expression at 12 hpi when compared to all other time points reaching levels of 3.2, the expression then decreased to 0.5 at 24 hpi and remained unchanged at 48 hpi (Figure 2d). The data showed similarity to the pyrosequencing data for this transcript.

The *profilin-like* gene was expressed constitutively at 1.9 prior to infection. Three hours after infection the transcript was significantly down-regulated to 0.6 (a 3 fold decrease) and remained unchanged at 6 hpi. There was a significant up-regulation from 6 hpi to12 hpi with expression peaking at 2.6, followed by a significant decrease to 0.9 at 24 hpi and remained unchanged at 48 hpi as opposed to 12 hpi (Figure 2e).

The MLO transmembrane protein encoding gene was constitutively expressed at 2.2 followed by a significant reduction at 6 hpi compared to the uninfected time point. At 12 hpi the tolerant

© University of Pretoria

rootstock responded with a significant increase to 5.7. Expression at 24 hpi was significantly down-regulated when compared to 12 hpi and reached a level of 1.5 at 48 hpi (Figure 2f). The *mlo expression* data was in agreement with the pyrosequencing data for this transcript.

The universal stress protein showed the maximum expression at 12 hpi but did not show a significant increase when compared to 0 and 6 hpi. An overall increase in expression was viewed until 12 hpi, followed by a significant down-regulation to 0.4 at 48 hpi compared to 12 hpi (Figure 2g).

Two genes encoding respectively for the thaumatin-like protein and the leucine rich repeat (LRR) resistance protein-like protein were constitutively expressed at approximately 0.8. Both genes showed no statistically significant change in regulation over the 48 hour time course, however similar to the majority of genes in this experiment, both genes showed the highest increase in expression at 12 hpi. The highest level of up-regulation achieved was 1.1 and 1.47 for the thaumatin-like gene and LRR protein-like gene respectively (Figure 2h & 2i).

The oxysterol binding protein generated a melting curve with multiple peaks. This was then visualised using gel electrophoresis and multiple banding was discovered. This primer set was then subjected to optimisation again but after failing the sequence was dropped from further analysis. It was omitted due to the nonspecific nature of the primers and generation of multiple melting curves.

3.4 Discussion

We have successfully produced the first molecular data for the avocado/*Phytophthora cinnamomi* interaction and quantified important defence-related genes in avocado roots in response to *P. cinnamomi*. The data generated from the quantification also confirmed the high-

© University of Pretoria

throughput sequencing data, demonstrating its use and importance in identification of important genes in non-model species. Transcripts were shown to be up-regulated at time points relative to the infection library that they were identified in.

Thaumatin, a PR5 protein associated with the SA pathway (Liu *et al.*, 2010; Wang *et al.*, 2010), was significantly upregulated at 48h in response to *P. cinnamomi* infection. The gene showed no changes in regulation during the first 6 hours after inoculation with a mycelial suspension. We suggest that the plant response is delayed due to the slower infection rate of mycelia as opposed to zoospores that are able to germinate within 2 hr of release from sporangia. At 12 and 24 hpi expression showed an insignificant but steady increase in response to the infection. The data suggests that thaumatin expression is positively influenced by the pathogen starting at 12 hpi and may play an important role in the qualitative defence response. PR5 is induced by biotic stress and further linked to increased pathogen resistance (Filippov *et al.*, 2005). García-Pineda *et al.* (2010) showed decreased root colonization in the avocado - *P. cinnamomi* system linked to SA. The significant up-regulation of thaumatin in the *P. cinammomi* tolerant R0.09 avocado rootstock could indicate the importance of the SA pathway in the early inhibition of the hemibiotroph *P. cinnamomi*. Hemibiotrophs have an initial biotrophic phase prior to becoming necrotrophs and PR5 gene activity in the SA-dependant pathway has been previosuly shown to be effective against biotrophs (Thomma *et al.*, 1998). Thaumatin also expresses a strong antifungal activity which inhibits mycelial growth (Liu *et al.*, 2010).

PsemI was highly expressed at 24 hpi in tolerant avocado roots infected with *P. cinnamomi.* The *PR10* gene was identified in the response of Douglas–fir infected with *Phellinus weirii* (Ekramoddoullah *et al.*, 2000). The authors showed that very high concentrations of *Pin m III* (*PsemI* gene homologue) was responsible for resistance to the rust pathogen *Cronartium ribicola*. This *PR-10* gene is valuable because it shows an up-regulation specifically by root

© University of Pretoria

pathogens and has also been used as a marker in screening for *P. weirii* resistance in Douglas-fir and could therefore be valuable in screening for PRR tolerance in avocado.

The gene encoding for the cytochrome P450-like TBP (Thylakoid binding protein) was the only transcript to be significantly induced by *P. cinnamomi* as early as 3 hpi. This enzyme features in oxidative metabolism and the production of ROS. This rapid response could be attributed to the universal nature of the protein in cell metabolism and growth. Additionally it has been reported to be involved in biotic and abiotic environmental responses as well as in the HR response to infection (Coram *et al.*, 2008; Fan *et al.*, 2011; Perrigault *et al.*, 2009; Thomas *et al.*, 2002).

The *metallothionein-like* gene showed a significant response to *P. cinnamomi* infection at 12 hours. Metallothioneins are cysteine-rich low molecular weight compounds that bind metal ions but their exact roles have not yet been identified (Whitelaw *et al.*, 1997). Metallothioneins have been shown to inhibit PCD and Fumonisin B1-induced root death in *Solanum lycopersicum* infected with *Agrobacterium rhizogenes* through interference of the ROS pathway. ROS accumulation was significantly reduced under metallothionein over-expression, validating its function in ROS scavenging (Harvey *et al.*, 2008). Significant induction of *metallothionein* in the highly tolerant avocado rootstock at 12 hpi implies that this protein may play a role in conferring disease tolerance to *P. cinnamomi* by scavenging ROS. ROS generation is indicative of the activity of the HR, which leads to cell death and is effective against biotrophic organisms (Morel & Dangl, 1997). This correlates with the rest of the data which suggests that the maximal plant defence response is at 12hours.

Two genes encoding for a profilin-like protein and a mlo–like transmembrane protein respectively, responded similarly to *P. cinnamomi* infection with a significant up-regulation at 12 hpi and a down-regulation at 24 hpi. Both these genes play a role in actin filament polarization (Opalski *et al.*, 2005; Sun *et al.*, 1995) and actin rearrangement has been observed in plant-

© University of Pretoria

fungus interactions with successful pathogen infection resulting in the suppression of the rearrangement (Kobayashi *et al.*, 1994; Kobayashi *et al.*, 1997).

It was observed in parsley, that upon infection with *Phytophthora infestans,* profilin localized to the site beneath the cell wall that is penetrated by the oomycetous appresorium (Schütz *et al.*, 2006) and either promotes or prevents actin polymerization in the actin cytoskeleton (Sun *et al.*, 1995). Cell wall thickening during fungal attack also involves the re-orientation of actin filaments as a defence response in order to prevent pathogen ingress (Schmelzer, 2002). The up-regulation of *profilin* in avocado roots suggests that profilin is being produced in response to *P. cinnamomi* penetration.

MLO, a transmembrane protein, relates extracellular signals into intra cellular responses. These proteins interact with calmodulin via a calmodulin binding domain. An increase in free calcium ion levels has been shown to be a response of plants to pathogen challenge (Kim *et al.*, 2002). The mlo transmembrane protein modulates actin cytoskeleton polarization in resistant barley in response to a biotrophic fungus - *Blumeria graminis* f. sp. *hordei* (Büschges *et al.*, 1997). Successful defence against pathogens results in cell wall strengthening and is correlated with increased actin accumulation at sites of attempted penetration (Opalski *et al.*, 2005).

Mlo gene expression in the highly tolerant avocado rootstock was pathogen responsive, resulting in a significant increase at 12 hpi and a down-regulation at 24 hpi. Our data correlates with the expression data of *mlo* in resistant barley plants infected with *Blumeria graminis* f. sp. *hordei* that showed an early 14 fold induction of *mlo* expression at 6 hpi followed by a down-regulation at 12, 24 and 48 hpi (Piffanelli *et al.*, 2002). We are however unable to comment on a second induction of the gene as our time course did not include time points beyond 48 hpi. Literature has highlighted opposing hypotheses for the exact role of *mlo* in plant-pathogen interactions. Barley wild-type and *mlo* mutants may modulate different mechanisms effective

© University of Pretoria

against biotrophic and necrotrophic fungi, in opposite directions (Kumar *et al.*, 2001; Piffanelli *et al.*, 2002). This study was not aimed at unravelling the role of *mlo* in avocado, but we have shown that the gene is induced upon infection by the hemibiotroph in a similar manner as a biotroph.

The up-regulation of the universal stress protein at 12 hpi indicates the plant's response to the stress of infection by *P. cinnamomi.* Universal stress proteins are mediated by ethylene (Sauter *et al.*, 2002), and our results may therefore implicate the involvement of the ethylene pathway in response to *P. cinnamomi,* a pathway that has shown activation in the *Arabidopsis-P. cinnamomi* interaction (Rookes *et al.*, 2008).

The thaumatin-like gene expression showed no significant response to the oomycete infection. This might be explained by the fact that it is usually induced by viral, bacterial and fungal infection (Liu *et al.*, 2010), and is believed to destroy fungal cell walls using a variety of enzymatic activities (Monteiro *et al.*, 2003). Since *P. cinnamomi* is an oomycete, it may not induce the gene.

The LRR (leucine-rich repeat) resistance protein-like gene also demonstrated no significant response to *P. cinnamomi.* Leucine-rich repeats are important in the transduction of plant internal and external signals (Tichtinsky *et al.*, 2003), they function as receptors to hormones and are also found to regulate abscisic acid (Osakabe *et al.*, 2005). Although resistance related LRR proteins have been found to interact specifically with other *Phytophthora* species, (*Solanum tuberosum-Phytophthora infestans* interaction) (Ballvora *et al.*, 2002), no such interaction has to our knowledge been described for *P. cinnamomi* and its hosts, specifically avocado. However, the identification of a LRR-like gene from *P. cinnamomi* infected tolerant avocado roots warrants further investigation.

© University of Pretoria

This study quantified an interesting set of genes regulated by the infection of the hemibiotroph *P. cinnamomi*. Defence genes included general defence-related transcripts (universal stress protein, metallothionein and thaumatin-like genes as well as cytochrome P450) and *Phytophthora* specific response genes such as (*thaumatin, PR-10* and the *LRR resistance protein-like* gene) (Ballvora *et al.*, 2002; Coram *et al.*, 2008; Ekramoddoullah *et al.*, 2000; Fan *et al.*, 2011; Filippov *et al.*, 2005; Liu *et al.*, 2010; Perrigault *et al.*, 2009; Sauter *et al.*, 2002; Thomas *et al.*, 2002; Wang *et al.*, 2010).

We inoculated avocado roots with a mycelial suspension that may explain a delay in the infection as opposed to infection by zoospores that are able to germinate and encyst within 2 hpi. Despite this delay five of the nine defence-related transcripts showed a significant early response to the pathogen between 3 and 12 hpi. Overall, we see a general trend in the increase of defence gene expression at 12hr. Based on this, we hypothesize that the plant generates a significant response against the pathogen at 12 hpi, possibly due to intracellular hyphal growth. This quick response generates tolerance against the pathogen.

3.5 References

Acheampong K. A., Akromah R., Ofori F. A., Takarama J. F., Saada D., Bitton I., Lavi U. (2008) Genetic characterization of Ghanaian avocados. *Journal of the American Society for Horticultural Science* **133**, 801-809.

Albert V., Soltis D., Carlson J., Farmerie W., Wall P. K., Ilut D., Solow T., Mueller L., Landherr L., Hu Y., Buzgo M., Kim S., Yoo M.-J., Frohlich M., Perl-Treves R., Schlarbaum S., Bliss B., Zhang X., Tanksley S., Oppenheimer D., Soltis P., Ma H., dePamphilis C.,

© University of Pretoria

Leebens-Mack J. (2005) Floral gene resources from basal angiosperms for comparative genomics research. *BMC Plant Biology* **5**, 5.

Ballvora A., Ercolano M. R., Weiß J., Meksem K., Bormann C. A., Oberhagemann P., Salamini F., Gebhardt C. (2002) The R1 gene for potato resistance to late blight (*Phytophthora infestans*) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. *The Plant Journal* **30**, 361-371.

Bekker T. F., Kaiser C., Labuschagne N. (2006) Efficacy of water soluble silicon against *Phytophthora cinnamomi* root rot of avocado: A progress report. *South African Avocado Growers' Association Yearbook* **29**, 58-62.

Borrone J. W., Schnell R. J., Violi H. A., Ploetz R. C. (2007) Seventy microsatellite markers from *Persea americana* Miller (avocado) expressed sequence tags. *Molecular Ecology Notes* **7**, 439-444.

Botha T., Kotze J. M. (1989) Exudates of avocado rootstocks and their possible role in resistance to *Phytophthora cinnamomi*. *South African Avocado Growers' Association Yearbook* **12**, 64-65.

Büschges R., Hollricher K., Panstruga R., Simons G., Wolter M., Frijters A., van Daelen R., van der Lee T., Diergaarde P., Groenendijk J., Töpsch S., Vos P., Salamini F., Schulze-Lefert P. (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. *Cell* **88**, 695-705.

Cass L. G., Kirven K. A., Christoffersen R. E. (1990) Isolation and characterization of a cellulase gene family member expressed during avocado fruit ripening. *Molecular and General Genetics* **223**, 76-86.

Chanderbali A. S., Albert V. A., Ashworth V. E. T. M., Clegg M. T., Litz R. E., Soltis D. E., Soltis P. S. (2008) *Persea americana* (avocado): bringing ancient flowers to fruit in the genomics era. *BioEssays* **30**, 386-396.

© University of Pretoria

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

Chen H., Ashworth V. E. T. M., Xu S., Clegg M. T. (2007) Quantitative genetic analysis of growth rate in avocado. *Journal of the American Society for Horticultural Science* **132**, 691-696.

Chernys J. T., Zeevaart J. A. D. (2000) Characterization of the 9-cis-epoxycarotenoid dioxygenase gene family and the regulation of abscisic acid biosynthesis in avocado. *Plant Physiology* **124**, 343-354.

Christoffersen R. E., Tucker M. L., Laties G. G. (1984) Cellulase gene expression in ripening avocado fruit: The accumulation of cellulase mRNA and protein as demonstrated by cDNA hybridization and immunodetection. *Plant Molecular Biology* **3**, 385-391.

Clegg M. T., Davis J. W. (1989) Molecular genetics of avocado. In: *California Avocado Society Yearbook*, pp. 59-61.

Clegg M. T., Henderson D., Durbin M. (1992) The development of molecular markers in avocado. *Proceedings of Second World Avocado Congress 1992 p. 573.*

Clegg M. T., Kobayashi M., Lin J. Z. (1999) The use of molecular markers in the managment and improvement of avocado (*Persea americana* Mill.). *Revista Chapingo Serie Horticultura* **5**, 227-231.

Coram T. E., Wang M., Chen X. (2008) Transcriptome analysis of the wheat–*Puccinia striiformis* f. sp. *tritici* interaction. *Molecular Plant Pathology* **9**, 157-169.

Davis J., Henderson D., Kobayashi M., Clegg M. T. (1998) Genealogical relationships among cultivated avocado as revealed through RFLP analyses. *Journal of Heredity* **89**, 319-323.

© University of Pretoria

Douhan G., Fuller E., McKee B., Pond E. (2011) Genetic diversity analysis of avocado (*Persea americana* Miller) rootstocks selected under greenhouse conditions for tolerance to phytophthora root rot caused by *Phytophthora cinnamomi*. *Euphytica* **182**, 209-217.

Ekramoddoullah A. K. M., Xueshu Y., Rona S., Arezoo Z., Doug T. (2000) Detection and seasonal expression pattern of a pathogenesis-related protein (PR-10) in Douglas-fir (*Pseudotsuga menziesii*) tissues. *Physiologia Plantarum* **110**, 240-247.

Fan F., Li X.-W., Wu Y.-M., Xia Z.-S., Li J.-J., Zhu W., Liu J.-X. (2011) Differential expression of expressed sequence tags in alfalfa roots under aluminum stress. *Acta Physiologiae Plantarum* **33**, 539-546.

Filippov A., Kuznetsova M., Rogozhin A., Spiglazova S., Smetanina T., Belousova M., Kamionskaya A., Skryabin K., Dolgov S. (2005) Increased resistance to late blight in transgenic potato expressing thaumatin II gene. *Ninth workshop of an European Network for development of an integrated control strategy of potato late blight*, 263-267.

Gachon C., Mingam A., Charrier B. (2004) Real-time PCR: what relevance to plant studies? *Journal of Experimental Botany* **55**, 1445-1454.

García-Pineda E., Benezer-Benezer M., Gutiérrez-Segundo A., Rangel-Sánchez G., Arreola-Cortés A., Castro-Mercado E. (2010) Regulation of defence responses in avocado roots infected with *Phytophthora cinnamomi* (Rands). *Plant and Soil* **331**, 45-56.

Hammond-Kosack K. E., Jones J. D. G. (1997) Plant disease resistance genes. *Annual Review of Plant Physiology and Plant Molecular Biology* **48**, 575-607.

Harvey J., Lincoln J., Gilchrist D. (2008) Programmed cell death suppression in transformed plant tissue by tomato cDNAs identified from an *Agrobacterium rhizogenes*-based functional screen. *Molecular Genetics and Genomics* **279**, 509-521.

© University of Pretoria

Hinch J. M., Clarke A. E. (1982) Callose formation in *Zea mays* as a response to infection with *Phytophthora cinnamomi*. *Physiological Plant Pathology* **21**, 113-124.

Kim M. C., Panstruga R., Elliott C., Muller J., Devoto A., Yoon H. W., Park H. C., Cho M. J., Schulze-Lefert P. (2002) Calmodulin interacts with MLO protein to regulate defence against mildew in barley. *Nature* **416**, 447-451.

Kobayashi I., Kobayashi Y., Hardham A. R. (1994) Dynamic reorganization of microtubules and microfilaments in flax cells during the resistance response to flax rust infection. *Planta* **195**, 237-247.

Kobayashi Y., Kobayashi I., Funaki Y., Fujimoto S., Takemoto T., Kunoh H. (1997) Dynamic reorganization of microfilaments and microtubules is necessary for the expression of non-host resistance in barley coleoptile cells. *The Plant Journal* **11**, 525-537.

Kumar J., Huckelhoven R., Beckhove U., Nagarajan S., Kogel K.-H. (2001) A Compromised Mlo pathway affects the response of barley to the necrotrophic fungus *Bipolaris sorokiniana* (teleomorph: *Cochliobolus sativus*) and its toxins. *Phytopathology* **91**, 127-133.

Kupke T., Hernández-Acosta P., Culiáñez-Macià F. A. (2003) 4′-phosphopantetheine and coenzyme A biosynthesis in plants. *Journal of Biological Chemistry* **278**, 38229-38237.

Liu J.-J., Sturrock R., Ekramoddoullah A. (2010) The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Reports* **29**, 419-436.

Monteiro S., Barakat M., Picarra-Pereira M. A., Teixeira A. R., Ferreira R. B. (2003) Osmotin and thaumatin from grape: a putative general defence mechanism against pathogenic fungi. *Phytopathology* **93**, 1505-1512.

Morel J. B., Dangl J. L. (1997) The hypersensitive response and the induction of cell death in plants. *Cell Death and Differentiation* **4**, 671-683.

© University of Pretoria

Opalski K. S., Schultheiss H., Kogel K.-H., Hückelhoven R. (2005) The receptor-like MLO protein and the RAC/ROP family G-protein RACB modulate actin reorganization in barley attacked by the biotrophic powdery mildew fungus *Blumeria graminis* f.sp. *hordei*. *The Plant Journal* **41**, 291-303.

Osakabe Y., Maruyama K., Seki M., Satou M., Shinozaki K., Yamaguchi-Shinozaki K. (2005) Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in *Arabidopsis*. *Plant Cell* **17**, 1105-1119.

Perrigault M., Tanguy A., Allam B. (2009) Identification and expression of differentially expressed genes in the hard clam, *Mercenaria mercenaria*, in response to quahog parasite unknown (QPX). *BMC Genomics* **10**, 377.

Phillips D., Grant B. R., Weste G. (1987) Histological changes in the roots of an avocado cultivar, Duke 7, infected with *Phytophthora cinnamomi*. *Phytopathology* **77**, 691-698.

Piffanelli P., Zhou F., Casais C., Orme J., Jarosch B., Schaffrath U., Collins N. C., Panstruga R., Schulze-Lefert P. (2002) The barley MLO modulator of defence and cell death is responsive to biotic and abiotic stress stimuli. *Plant Physiology* **129**, 1076-1085.

Rookes J. E., Wright M. L., Cahill D. M. (2008) Elucidation of defence responses and signalling pathways induced in *Arabidopsis thaliana* following challenge with *Phytophthora cinnamomi*. *Physiological and Molecular Plant Pathology* **72**, 151-161.

Sánchez-Pérez J., Jaimes-Lara M., Salgado-Garciglia R., López-Meza J. (2009) Root extracts from Mexican avocado *Persea americana* var. *drymifolia* inhibit the mycelial growth of the oomycete *Phytophthora cinnamomi*. *European Journal of Plant Pathology* **124**, 595-601.

Sauter M., Rzewuski G., Marwedel T., Lorbiecke R. (2002) The novel ethylene-regulated gene *OsUsp1* from rice encodes a member of a plant protein family related to prokaryotic universal stress proteins. *Journal of Experimental Botany* **53**, 2325-2331.

© University of Pretoria

Schmelzer E. (2002) Cell polarization, a crucial process in fungal defence. *Trends in Plant Science* **7**, 411-415.

Schütz I., Gus-Mayer S., Schmelzer E. (2006) Profilin and Rop GTPases are localized at infection sites of plant cells. *Protoplasma* **227**, 229-235.

Sun H.-Q., Kwiatkowska K., Yin H. L. (1995) Actin monomer binding proteins. *Current Opinion in Cell Biology* **7**, 102-110.

Thomas S. W., Glaring M. A., Rasmussen S. W., Kinane J. T., Oliver R. P. (2002) Transcript profiling in the barley mildew pathogen *Blumeria graminis* by serial analysis of gene expression (SAGE). *Molecular Plant-Microbe Interactions* **15**, 847-856

Thomma B. P. H. J., Eggermont K., Penninckx I. A. M. A., Mauch-Mani B., Vogelsang R., Cammue B. P. A., Broekaert W. F. (1998) Separate jasmonate-dependent and salicylate-dependent defence-response pathways in *Arabidopsis* are essential for resistance to distinct microbial pathogens. *Proceedings of the National Academy of Sciences* **95**, 15107-15111.

Tichtinsky G., Vanoosthuyse V., Cock J. M., Gaude T. (2003) Making inroads into plant receptor kinase signalling pathways. *Trends in Plant Science* **8**, 231-237.

Wang X., Tang C., Deng L., Cai G., Liu X., Liu B., Han Q., Buchenauer H., Wei G., Han D., Huang L., Kang Z. (2010) Characterization of a pathogenesis-related thaumatin-like protein gene TaPR5 from wheat induced by stripe rust fungus. *Physiologia Plantarum* **139**, 27-38.

Whitelaw C. A., Le Huquet J. A., Thurman D. A., Tomsett A. B. (1997) The isolation and characterisation of type II metallothionein-like genes from tomato (*Lycopersicon esculentum* L.). *Plant Molecular Biology* **33**, 503-511.

Zentmyer G. A. (1984) Avocado diseases. *Tropical Pest Managment* **30**, 388-400.

© University of Pretoria

3.6 Figures & Tables

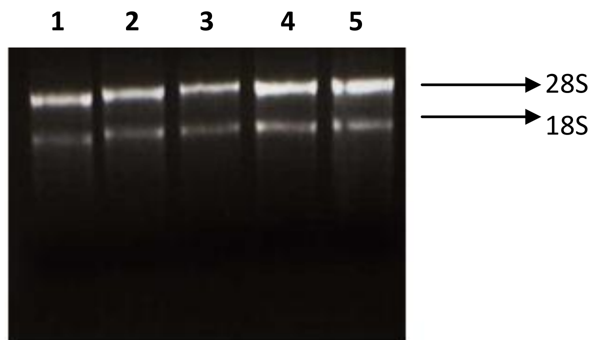
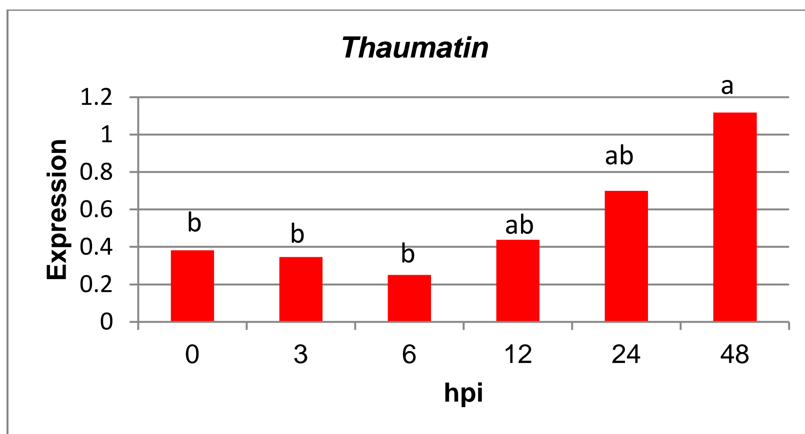


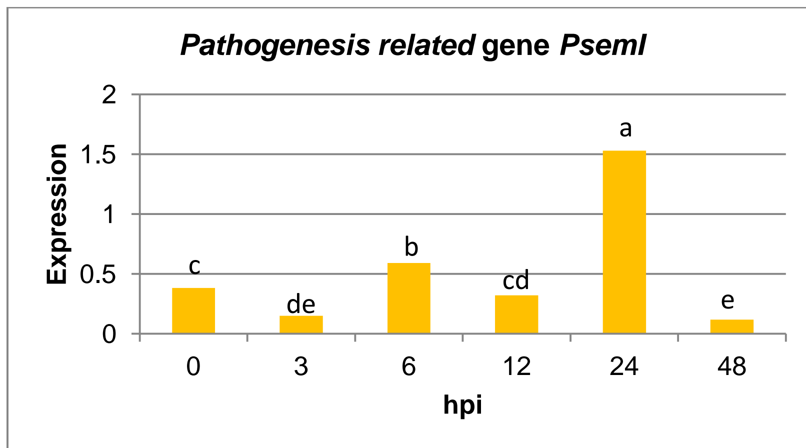
Figure 1. RNA extraction of *Phytophthora cinnamomi* infected tolerant R0.09 avocado roots. Lanes 1-5; RNA extracted from roots of avocado rootstock, R0.09, at various time points with two distinct 28S and 18S rRNA bands. Lane 1: 0 hour RNA isolation, lane 2: 6 hour RNA isolation, lane 3:12 hour RNA isolation, lane 4:24 hour RNA isolation and lane 5: 48 hour RNA isolation.

2a)

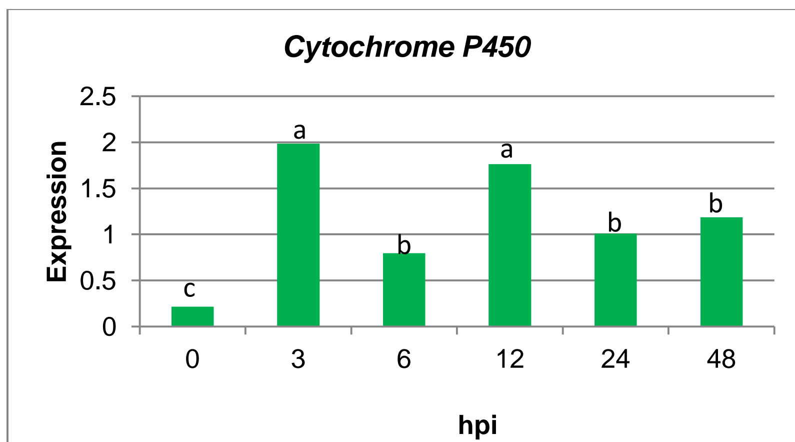


© University of Pretoria

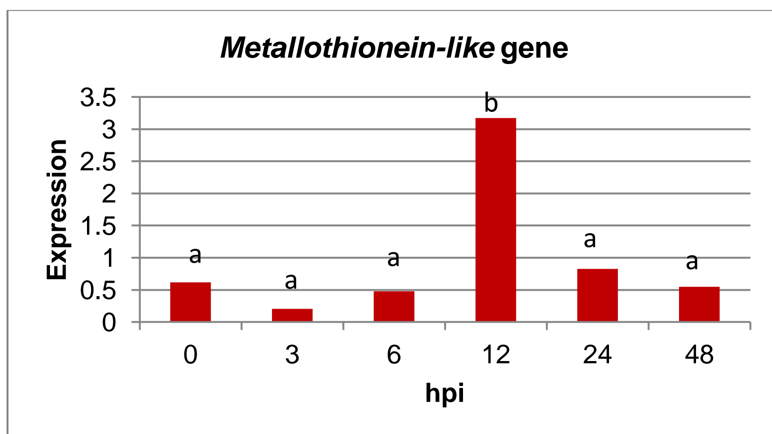
2b)



2c)

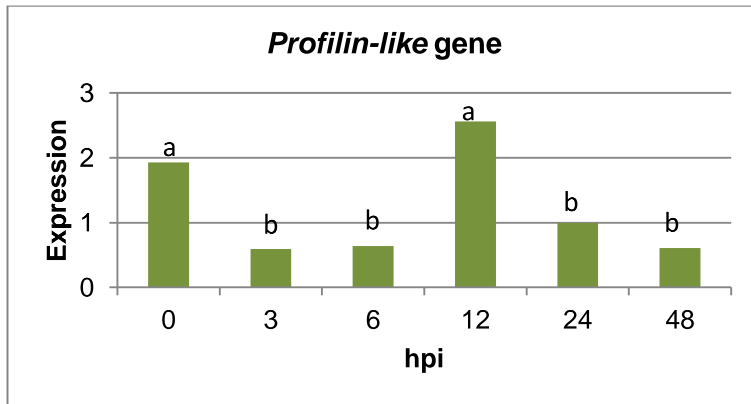


2d)

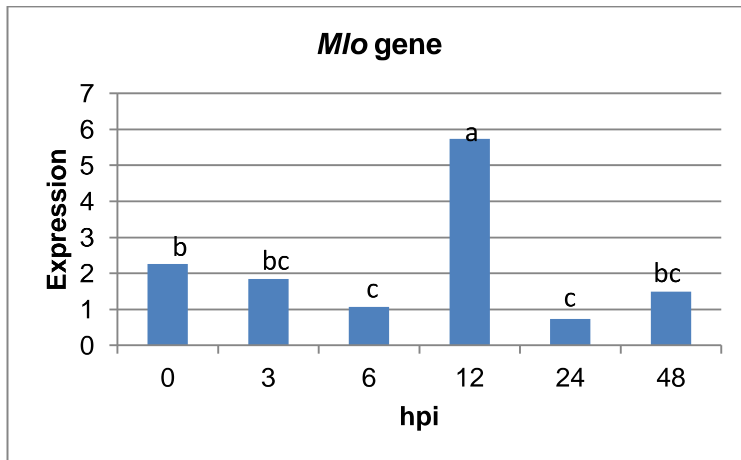


© University of Pretoria

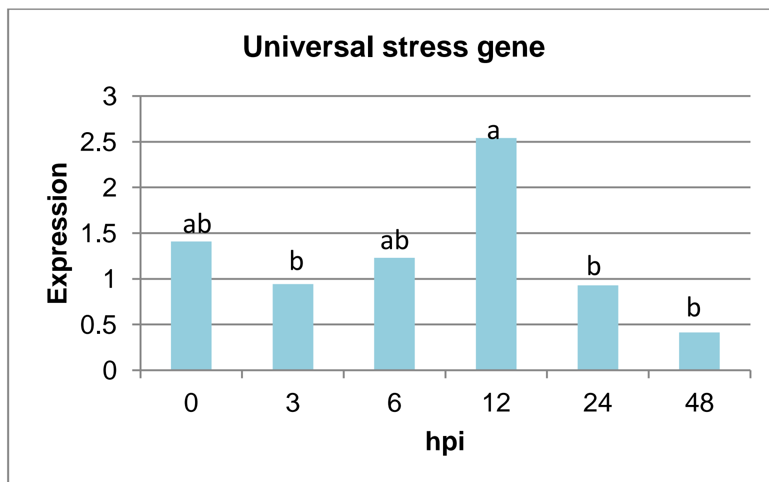
2e)



2f)

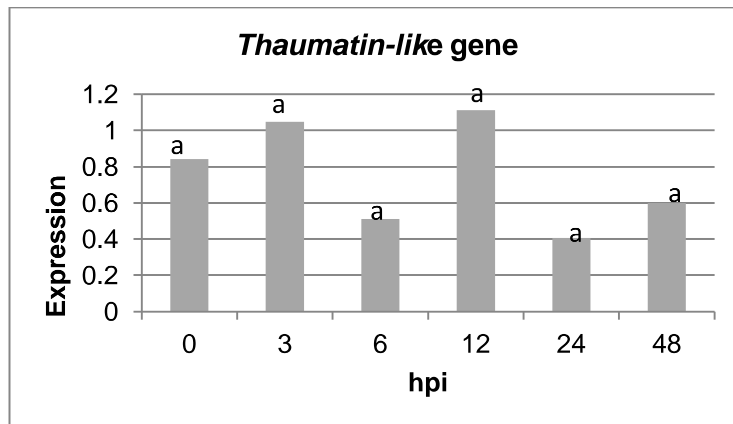


2g)



© University of Pretoria

2h)



2i)

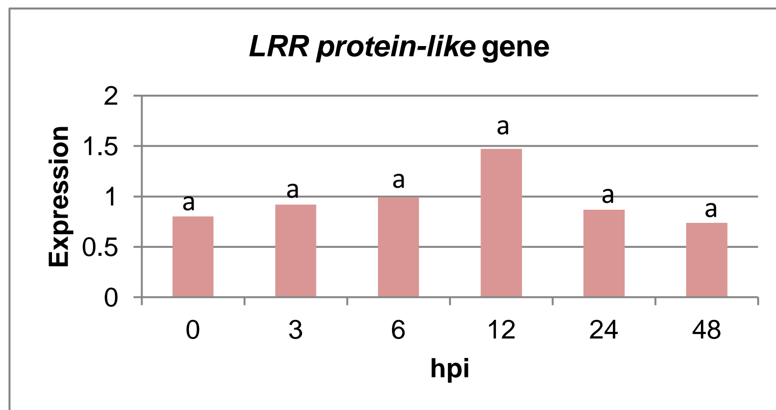


Figure 2. Gene expression of a tolerant avocado rootstock, infected with *Phytophthora cinnamomi.*

Expression analysis was conducted at 0, 3, 6, 12, 24 & 48 hpi (hours post infection) with 0 hr being the uninfected control. The data was normalized using two reference genes- actin and 18S. Expression analysis was performed in triplicate on three biological replicates. Statistical differences of expression levels are indicated with letters (ab, a, cd) above time points. Expression levels with different letters are significantly different (Student's t test ($P < 0.05$)). If two time points have the same letter(s) above then they are not significantly different **(a)** *Thaumatin*. **(b)** Pathogenesis-related protein *pseml* (PR10). **(c)** *Cytochrome P450.* **(d)**

© University of Pretoria

Metallothionein-like gene. **(e)** *Profilin-like* gene. **(f)** *MLO* transmembrane gene. **(g)** The universal stress protein. **(h)** The *thaumatin-like* gene. **(i)** The LRR resistance protein-like protein.

© University of Pretoria

Table 1. Defence-related genes identified in R0.09 avocado rootstock infected with *Phytophthora cinnamomi.*

Sequence	Defence-related genes	BLASTX organism identity
Contig0011 (infected library 2)	*Thaumatin*	*Vitis riparia*
Contig0020 (infected library 1)	*LRR resistance-like*	*Gossypium hirsutum*
Contig0064 (infected library 1)	*Metallothionein-like*	*Persea americana*
Contig0073 (infected library 2)	*Oxysterol-binding*	*Solanum tuberosum*
Contig0083 (infected library 2)	*Thaumatin-like*	*Arabidopsis thaliana*
Contig0088 (infected library 1)	*Mlo transmembrane*	*Zea mays*
Contig0106 (infected library 1)	*Pathogenesis related PsemI*	*Pseudotsuga menziesii*
Contig0108 (infected library 2)	*Proflin-like*	*Cinnamomum camphora*
Contig0175 (infected library 2)	*Putative universal stress*	*Cicer arietinum*
Contig0187 (infected library 2)	*Cytochrome P450 like*	*Nicotiana tabacum*

© University of Pretoria

Table 2. Primer sequences of selected putative avocado defence-related genes from *Phytophthora cinnamomi* infected avocado roots. Sequences were chosen from the library 1 and library 2 based on differential expressions. Two reference genes - actin and 18s were included.

Sequence ID	Fwd primer (5'-3')	Rev primer (5'-3')	Product length (bp)	Annealing °C
Thaumatin	CACCCTGTAGTTCACTCC	CCAGATGCTTACAGTTACC	75	58.5
LRR resistance PLP	GACATTCTTATAGCCATC	ATAAACAATCTGATTTTG	135	56
Metallothionein like protein type 2	AGTCTTCATCCCTAATACATATCCC	GTTTGTGCGTGTCTGGTTTC	76	58.5
Thaumatin-like protein	AAGCAGTCCTCAAGGTTC	TTTCCGTTAGTGTCAAAGC	79	65
MLO protein	TCGTGGATGGAAGGAGTG	ATGGGCAAATCTAAATCTTGTTG	85	58.5
Pathogenesis-related protein PsemI	GAAGATGGAGTACAAATAC	CACCTTGATGTGATAAAC	82	58.5
Proflin-like protein	TTCGGTATCTATGATGAG	ACGATATGACATTCAATAG	110	58.5
Putative universal stress protein	GACATTCTTATAGCCATC	ATAAACAATCTGATTTTG	135	56
Cytochrome P450 like TBP	GTCAAAGTGAAGAAATTC	AATCTCGTTAATCCATTC	119	58.5
18S	GTCAAAGTGAAGAAATTC	AATCTCGTTAATCCATTC		58.5
Actin	GAATCTGGACCATCTATTG	TACCAACCAAACCAAATC	114	58.5

© University of Pretoria

Table 3. Similarities between pyrosequencing data and gene expression profiles of defence-related genes. All the genes chosen for expression profiling from the tolerant avocado rootstock infected with *Phytophthora cinnamomi* showed the highest expression from a time point related to the cDNA library of their identification except the pathogenesis-related protein, profilin-like protein and the universal stress protein.

Defence-related gene	GenBank accession number	cDNA Library	Max qRT-PCR expression	Similarities between 454 and qRT-PCR
Thaumatin	JO840464	Library 2	48 hpi	yes
LRR resistance PLP	JO840460	Library 1	12 hpi	yes
Metallothionein like protein	JO840461	Uninfected, Library 1	12 hpi	yes
Thaumatin-like protein	JO840465	Library 2	12 hpi	yes
Seven transmembrane protein MI0	JO840462	Library 1	12 hpi	yes
Pathogenesis-related protein PsemI	JO840463	Library 1	24 hpi	no
Proflin-like protein	JO840466	Library 2	12 hpi	no
Putative universal stress protein	JO840467	Library 2	12 hpi	no
Cytochrome P450 like TBP	JO840468	Library 1 and Library 2	3/ 12 hpi	yes

© University of Pretoria

CHAPTER 4

General discussion

© University of Pretoria

Avocados are an important crop that generates billions of dollars in revenue in 1st world countries. Aside from their economic importance they are also high in monounsaturated fat and selected vitamins. Their production is constantly under threat by the pathogen *Phytophthora cinnamomi* which causes Phytophthora root rot (PRR). The current pathogen control methods are threatening to become ineffective and the lack of understanding of defence strategies against the pathogen necessitates investigation into the molecular structure of tolerance.

Understanding defence against *P. cinnamomi* in avocado is an important for the development of disease resistant plants. Root pathogens and their interactions with hosts are poorly understood. Despite the importance of both the avocado crop and the effect *Phytophthora* has on its cultivation and although the plant-pathogen interaction has been documented for over 60 years there is a lack of molecular knowledge underpinning our understanding of defence strategies against the pathogen. The study of gene expression enables us to determine what may constitute a tolerant phenotype in avocado rootstocks. This study was aimed at arming the avocado research community with a selection of defence-related ESTs generated from the tolerant rootstock R0.09 so that further studies could be conducted in the fight against *P. cinnamomi*. The generation of these ESTs marked the beginning of the molecular dissection of the avocado transcriptome. The goals achieved were; the generation of sequence data for the root of the avocado, the identification of defence related genes in response to *P. cinnamomi* and the quantification of defence gene expression.

The lack of genetic information available for avocado necessitated the use of a high-throughput sequencing platform to generate ESTs. This method would be the most cost effective, rapid and error free strategy in non-model organisms such as avocado. The 454 GS-FLX pyrosequencing platform was used to generate EST data from avocado. Even though a reference genome is not available the longer read lengths enable the successful assembly of contigs. Since our aim was

© University of Pretoria

to investigate general transcript expression over the entire root transcriptome and not identify only differentially expressed genes, it did not include a subtraction step. However, if selective subtractive hybridization (SSH) is combined with high-throughput sequencing then a large number of differentially expressed genes can be identified. Combining the two technologies have demonstrated significant advantages in identifying differentially expressed genes. The amount of data that can be produced using high-throughput sequencing at reduced costs, time and labour minimises the need for other methods to enrich for low copy number transcripts.

Sufficient ESTs were generated to identify genes involved in various functions. Due to the high sensitivity of sequencing, transcriptome studies identify many transcripts that have not yet been characterised and many that have unknown functions even when annotated using a database such as Gene ontology. The 2 Mb of data generated gave a broad representation of gene expression and assembled into 371 contigs of which 367 ESTs are novel for *P. americana* and had not previously been identified. The top ten functional groupings according to the GO classification revealed that 44.5 % of assembled contigs were represented by unknown functions followed by the functional groups of ‘other’, ‘cellular components’, ‘biological processes’, ‘stress responses’, ‘ribosome structure’, ‘cell wall related’, ‘protein binding’, ‘mitochondrion and ‘ATP-binding’. Of specific interest were the transcripts that showed possible involvement in the defence pathways linked to *P. cinnamomi*. The lack of EST and genome sequence data for avocado in general, specifically rootstocks, also accounts for the high frequency of unknown functions observed.

Avocado is an economically important basal angiosperm, a group of plants that has very little sequence data available on public databases. The lack of homology between our sequence database and the NCBI database demonstrates the deficiency of avocado sequence data. Only 1 % (4/371) of the sequenced contigs showed homology to *P. americana*, and represented only

© University of Pretoria

two genes- metallothionein-like protein type 2 and fructose- bisphosphate aldolase. Interestingly 20-30% of the contigs generated showed similarity to grapevine (*V. vinifera*) when the BLAST sequence homologies were analysed. The sequence similarities were not restricted to any specific functional class, thus precluding the homology observed being attributed to highly conserved genes.

The first set of pathogen linked ESTs were founded in this study. These represented both previously identified *Phytophthora* specific as well as general defence-related genes. Interesting defence ESTs included *thaumatin, metallothionein, a PR10* pathogenesis-related protein, a mlo transmembrane protein and *profilin*. Nine genes were quantified with qRT-PCR to elucidate the early gene response of a tolerant avocado rootstock infected with *P. cinnamomi*. The aim of the expression analysis was to quantify interesting defence-related genes but also to validate the use of pyrosequencing in the non-model avocado plant.

The transcripts chosen for quantification represent both broad and specific defence characteristics. Quantification of the defence genes revealed that expression was varied and although some transcripts demonstrated a steady increase over the tested time points, others demonstrated an increase followed by a decrease or a gradual overall reduction across time points. The study indicated 66% of the genes validated the high-throughput sequencing data verifying the techniques suitability.

The application of high-throughput sequencing methods enables a foundation for hypothesis-generation for crop improvement. Based on this first set of transcriptome data we hypothesize that the tolerance of the rootstock in this study is most likely polygenic and based on the early detection of *P. cinnamomi* followed by a response that included ROS and cell-wall strengthening. We inoculated avocado roots with a mycelial suspension that may explain a delay in the infection as opposed to infection by zoospores that are able to germinate and

© University of Pretoria

encyst within 2 hpi. Despite this delay five of the nine defence-related transcripts showed a significant early response to the pathogen between 3 and 12 hpi. Our current hypothesis is that the critical time of defence activation is between 3 and 12 hpi for pathogen recognition along with a maximal plant response being generated at 12 hours.

There were some short comings in the study. A full set of non-inoculated control plants for each time point was omitted purely based on resource constraints. We designed the experiment as best we could, with the small number of plants that were available. Three biological replicates were used per time point to avoid anomalies and provide meaningful expression data. It would have been valuable to compare the expression profiles of the defence genes across different rootstocks or on a known susceptible rootstock infected with *Pc.* This would allow validation of the potential candidacy of those genes as molecular markers for breeding and selection programs. Currently this study is being followed up by a study comparing the defence response of 5 rootstocks using the genes identified, comments on the use of these genes as molecular markers can be assessed once it is complete. Minimal sequence data was generated due to using a single sequencing lane. If more sequencing was done more transcripts would have been annotated and perhaps more defence-related genes found. More time points should have been used to get a better picture of the expression patterns of the avocado response. Finally a combination of different methods of sequencing needs to be employed (time and cost dependent) to generate more depth and better assembly of the data.

We have successfully produced the first molecular data for the avocado/*Phytophthora cinnamomi* interaction and believe that this data will contribute to the understanding of host defence against this devastating pathogen thereby aiding in the selection of tolerant avocado rootstocks. This research is important for the development of rootstock breeding, which has been indicated as a key aspect of improving and sustaining the avocado industry very early in

© University of Pretoria

the crops commercialization. The data generated will form a basis for further transcriptomic work and gene expression studies to be done. It will help in the elucidation of the transcriptome, identification of markers for improved rootstock breeding and screening and assist in microarray analysis for avocado root expression to become a reality. New chemical controls can be developed from studying how avocado interacts with the pathogen. The gene discovery initiated by this study will allow biochemical studies to be carried out and can be studied further in other rootstocks and is the beginning of identifying the genes that play a role in the host-pathogen interaction. The ESTs generated for avocado in this study have already been useful in providing transcript data to address other research questions such as the isolation of the full length cDNA sequence of thaumatin and of a putative *NPR1* gene homolog as well as investigating the hypoxia response in avocado roots after flooding. These ESTs are valuable for use in gene discovery for an organism that has very little or no genome data available. This research is critical in understanding the infection and response processes in the interaction because it cannot be conclusively answered by studies on non-hosts. Due to the importance of avocado as a fruit crop and the insight that the basal angiosperm provides to other magnolids or phylogenetically related plants, it is important that we expand on the knowledge base of molecular information. The genetic information generated and defence-related genes identified in this study will be an invaluable resource in the current agricultural and biotechnological landscape, bringing the avocado industry a step closer to elucidating the defence response against *P. cinnamomi*.

© University of Pretoria