ANALYSIS OF THE EARLY EVENTS IN THE INTERACTION BETWEEN *VENTURIA INAEQUALIS* AND THE SUSCEPTIBLE GOLDEN DELICIOUS APPLE (*MALUS* X *DOMESTICA* BORKH.).

LIZEX HOLLENBACH HERMANUS HÜSSELMANN

A thesis submitted in partial fulfilment of the requirements for the degree of *Philosophiae Doctorae* in the Faculty of Science, University of the Western Cape.

Supervisor: Prof Bongani K. Ndimba

Co-supervisor: Dr D. Jasper G. Rees

October 2014

TABLE OF CONTENTS

ABSTRACT

Analysis of the early events in the interaction between *Venturia inaequalis* and the susceptible Golden Delicious apple (*Malus* x *domestica* Borkh.)

PhD thesis, Department of Biotechnology, Faculty of Science, University of the Western Cape.

Lizex Hollenbach Hermanus Hüsselmann

Apple (*Malus* x *domestica*) production in the Western Cape, South Africa, is one of the major contributors to the gross domestic product (GDP) of the region. The production of apples is affected by a number of diseases. One of the economically important diseases is apple scab that is caused by the pathogenic fungus, *Venturia inaequalis*. Research to introduce disease resistance ranges from traditional plant breeding through to genetic manipulation. Parallel disease management regimes are also implemented to combat the disease, however, such strategies are increasingly becoming more ineffective since some fungal strains have become resistant to fungicides. The recently sequenced apple genome has opened the door to study the plant pathogen interaction at a molecular level.

This study reports on proteomic and transcriptomic analyses of apple seedlings infected with *Venturia inaequalis*. In the proteomic analysis, two-dimensional gel electrophoresis (2-DE) in combination with mass spectrometry (MS) was used to separate, visualise and identify apple leaf proteins extracted from infected and uninfected apple seedlings. Using

Melanie™ 2-DE Gel Analysis Software version 7.0 (Genebio, Geneva, Switzerland), a comparative analysis of leaf proteome expression patterns between the uninfected and infected apple leaves were conducted. The results indicated proteins with similar expression profiles as well as qualitative and quantitative differences between the two leaf proteomes. Thirty proteins from the apple leaf proteome were identified as differentially expressed. These were selected for analysis using a combination of MALDI-TOF and MALDI-TOF-TOF MS, followed by database searching. Of these spots, 28 were positively identified with known functions in photosynthesis and carbon metabolism (61%), protein destination and storage (11%), as well as those involved in redox/response to stress, followed by proteins involved in protein synthesis and disease/defence (7%), nucleotide and transport (3%).

RNA-Seq was used to identify differentially expressed genes in response to the fungal infection over five time points namely Day 0, 2, 4, 8 and 12. cDNA libraries were constructed, sequenced using Illumina HiScan SQ™ and MiSeq™ instruments. Nucleotide reads were analysed by aligning it to the apple genome using TopHat splice-aware aligner software, followed by analysis with limma/voom and edgeR, R statistical packages for finding differentially expressed genes. These results showed that 398 genes were differentially expressed in response to fungal infection over the five time points. These mapped to 1164 transcripts in the apple transcripts database, which were submitted to BLAST2GO. Eighty-six percent of the genes obtained a BLAST hit to which 77% of the BLAST hits were assigned GO terms. These were classed into three ontology categories i.e. biological processes, molecular function and cellular components. By

focussing on the host responsive genes, modulation of genes involved in signal perception, transcription, stress/detoxification, defence related proteins, transport and secondary metabolites have been observed.

A comparative analysis was performed between the Day 4 proteomic and Day 4 transcriptomic data. In the infected and uninfected apple leaf proteome of Day 4, we found 9 proteins responsive to fungal infection were up-regulated. From the transcriptome data of Day 4, 162 genes were extracted, which mapped to 395 transcripts in the apple transcripts. These were submitted to BLAST2GO for functional annotation. Proteins encoded by the up-regulated transcripts were functionally categorised. Pathways affected by the up-regulated genes are carbon metabolism, protein synthesis, defence, redox/response to stress. Up-regulated genes were involved in signal perception, transcription factors, stress/detoxification, defence related proteins, disease resistance proteins, transport and secondary metabolites. We found that the same pathways including energy, disease/defence and redox/response to stress were affected for the comparative analysis. The results of this study can be used as a starting point for targeting host responsive genes in genetic manipulation of apple cultivars.

KEYWORDS

Fruit crops, apple proteomics, apple transcriptomics, RNA-Seq technology, next generation sequencing, MALDI-TOF MS, MALDI-TOF-TOF MS, edgeR and limma/voom statistical packages, protein identification, host-pathogen interaction study.

DECLARATION

I declare that, **Analysis of the early events in the interaction between *Venturia inaequalis* and the susceptible Golden Delicious apple (*Malus* x *domestica* Borkh.)**, is my own work that has not been submitted for any degree or examination in any other university and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Lizex Hollenbach Hermanus Hüsselmann October 2014

Signed:…….

ACKNOWLEDGEMENTS

I would like to especially thank the following people and institutions:

- My Lord and Saviour Jesus Christ for the health, patience, perseverance and strength to fulfil my dreams.
- My mother, my wife Esther, my brothers and sisters for your love, support, inspiration and motivation to do what was required to finish this degree.
- Prof Jasper Rees, for the opportunity to become part of the plant genetics research group at the University of the Western Cape and for all the guidance during my PhD studies. Thank you for all the opportunities to present my work nationally as well as internationally and all the associated opportunities to travel and attend workshops related to my field of interest. Thank you for the financial support that made it possible for me to further my academic career.
- Prof Bongani Ndimba, for all your help and advice in the proteomics section of the project and as well as the administrative role you played during this process.
- All the members of the Fruit tree genetics group, formerly at the University of the Western Cape, for your friendship, support and helpful discussions. A special thank you to Dr Devon Ryan (German Centre for Neurodegenerative Diseases (DZNE), Bonn, Germany) and Dr Stanley Mbandi (SANBI) for your assistance in writing the bioinformatics scripts and useful discussions.
- A special word of thanks to Prof Ludidi Ndiko for critically reviewing the thesis.
- Support staff in the Department of Biotechnology especially Dr Mervin Meyer for advice and moral support at the University of the Western Cape and Agricultural Research Council at Infruitec.

LIST OF FIGURES

LIST OF TABLES

APPENDIX A

APPENDIX B

APPENDIX C

ABBREVIATIONS

A	adenine
AFLP	amplified fragment length polymorphism
APS	ammonium persulphate
BLAST	basic local alignment search tool
bp	base pair
C	cytosine
CBB	coomasie brilliant blue
CCD	charge coupled device
cDNA	complementary DNA
CHS	chalcone synthase
CO_2	carbon dioxide
CTAB	cetyltrimethylammonium-bromide
cv	cultivar
dpi	days post inoculation
2D	second dimension
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EC	evidence code
EDTA	ethylene diamine tetra acetic acid
ESI	electrospray ionization
EST	expressed sequence tags
FPKM	fragment per kilobase pair of exon model per million fragments mapped
G	guanine
gDNA	genomic DNA
GDR	genome database for Rosaceae
GDP	gross domestic product

GFP	green fluorescent protein
GO	gene ontology
hpi	hour post inoculation
hps	minimal alignment length
IASAMA	Istituto Agrario San Michelle all'Adige
IPG	immobilized pH gradient
JA	jasmonates
Kb	kilobase
KEGG	kyoto encyclopaedia of genes and genomes
LB	luria broth
LCM	laser capture micro-dissection
M	molar
MALDI	matrix-assisted laser desorption ionization
Mb	mega base
Mg^{+2}	magnesium cation
MgCl	magnesium chloride
$mg.ml^{-1}$	milligram per millilitre
min	minutes
miRNA	micro RNA
ml	milliliter
mM	millimolar
MPSS	massively parallel signature sequencing
mRNA	messenger ribonucleic acid
m/v	mass per volume
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MT	metric ton
NaCl	sodium chloride
cRNA	coding RNA
ng	nanogram
nr	non-redundant

PAGE	polyacrylamide gel electrophoresis
pI	isoelectric point
PAL	phenylalanine ammonia-lyase
piRNA	piwi-interacting RNA
PCR	polymerase chain reaction
pH	power of hydrogen
RAPD	random amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RH	relative humidity
RNA	ribonucleic acid
rRNA	ribosomal RNA
rpm	revolutions per minute
S	south
SA	salicylic acid
SAGE	serial analysis of gene expression
siRNA	short interfering RNA
snoRNA	small nuclear RNA
tRNA	transfer RNA
T	thymidine
TCA	trichloroacetic acid
TBE	tris, boric acid, EDTA
TEMED	N,N,N,N-tetramethylethylenediamine
TM	Trade Mark
TOF	time of flight
Tris-HCl	tris[hydroxymethyl]aminomethane hydrochloric acid
TSP	total soluble protein
UK	united kingdom of England
µg	microgram
µl	microliter
µM	micromolar
US	university of Stellenbosch

UV	ultraviolet
V	volts
v/v	volume per volume
w/v	weight per volume

Chapter One – Literature Review

1 Introduction

1.1 Economic importance of apple

Apples constitute the largest deciduous fruit crop in the Western Cape province of South Africa. This fruit is ranked fourth in world fruit production behind bananas, oranges, and grapes (http://faostat.fao.org). Apples are economically important for the many ways in which they can be consumed such as fresh or dried fruit, juices or alcoholic beverages and pulp for candies. They are versatile and can be marketed fresh or processed. Total production for the period 2011/2012 was 815 000 metric ton (MT) of which fresh apple consumption is 220 000 MT, for processing 165 000 MT and the export market 430 000 MT (Pickelsimer, 2013). Apples are exported mainly to Europe with the United Kingdom of England (UK) its biggest market, 98 000 MT (Pickelsimer, 2013). Additional demand from non-traditional African and Middle East markets were experienced (Pickelsimer, 2013). Due to the huge market for apples, a substantial amount in revenue is contributed to the Western Cape gross domestic product (GDP) (Siphugu, 2011). The apple industry therefore, plays a crucial role in the economy of South Africa and creates employment opportunities for permanent as well as seasonal workers.

1.2 Origin and spread of apples

Apple, together with pear, plum peach, cherry strawberry, raspberry and blackberry belong to the Rosaceae family. Apples and pears amongst other have been classified into the subfamily pomoideae commonly known as pome fruits. The reason for this classification is that their fruits consist of two to five carpels, which are enclosed in a fleshy covering. The cultivated apple in particular is designated by the scientific name *Malus* x *domestica*, believed to be the result of an interspecific hybridisation event (Korban and Skirvin, 1984). Already in 1930 Nicolai Vavilov speculated that the wild apple of Turkestan and its close relatives are the ancestors of the domesticated apple, which could be traced back to Almaty (Father of Apples; Kazakhstan; Harris *et al.*, 2002). The progenitor of apple has previously been considered to be *Malus sieversii* based on morphological, molecular and historical evidence (Cornille *et al.*, 2012). *Malus sieversii* originated from the Heavenly Mountains (Tien Shan) at the boundary between western China and the former Soviet Union and these forests were identified as the geographic area where the apple was first domesticated (Janick *et al.*, 1996; Cornille *et al.*, 2012). Forseline *et al.* (1994) confirmed that *M. sieversii* is very diverse and also that it contains all the qualities present in *Malus* x *domestica*. However more recently, new insight on the origin of apple has been brought forward. Cornille *et al.* (2012) postulated that multiple species have contributed to the genetic makeup of domesticated apples. One species namely *Malus sylvestris* in particular was a major secondary contributor. Furthermore, bidirectional gene flow between the domesticated apple and the European crabapple resulted in *M. domestica* being more closely related to *M. sylvestris* than to *M. sieversii* (Cornille *et al.*, 2012).

The spread of apples started as early as 5000-8000 years ago when Central Asia was crossed by the famous Silk Roads stretching from Rome in Italy through Samarkand in Uzbekistan to Luoyang in China (Fig. 1.1) (Wood, 2003). Apple cultivation likely began in the region between the Caspian and Black seas, and it had reached the Middle East by 3,000 years ago (Hancock, 2004). The Romans skilfully introduced and spread apple through their grafting techniques across the European and Mediterranean areas. European settlers have transported it since that time to the rest of the world. Today, apple is grown in all temperate regions of the world, including South Africa. The challenge today, however, is still to produce fruit free from pests and diseases.

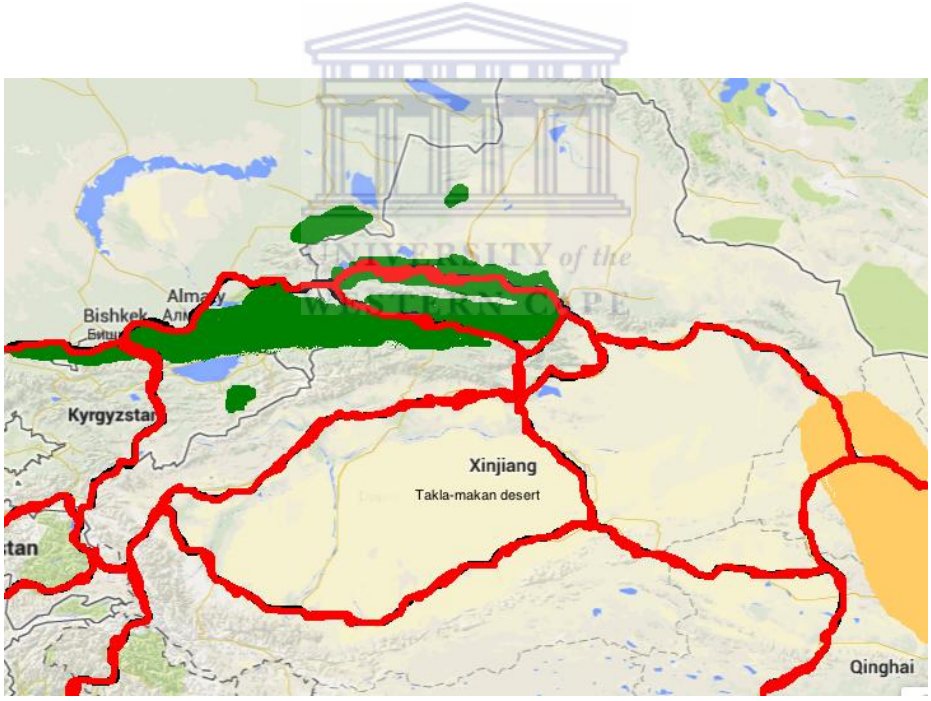


Figure 1.1 Distribution map of ancient apples. Maps showing distribution of ancient apples via the silk roads from Kazakstan to China (Adjusted from Harris *et al.*, 2002). Orange – region of high concentration of 'ancient' apples, Dark-green – Pre-history surmised distribution of *Malus sieversii*, Red – Silk roads

1.3 Diseases and pests

The market is very dynamic and the quality criterion varies depending on the targeted clientele. Therefore, a breeder's concept of quality should be in alignment with that of the intended market. Apple prices reflect the market's perception of acceptable appearance (size, colour, shape and blemish-free) and quality (taste and firmness). All these are subject to being free from pests and diseases. Pests include apple maggots, plum curculio and codling moth, which are some of the most common pests that attack apple trees. Furthermore, there are a number of important diseases of apples such as powdery mildew, fireblight, woolly apple aphid and apple scab that cause millions of Rands in damages to apple orchards.

1.3.1 Apple scab

Venturia inaequalis (Cooke) Winter, the causal agent of apple scab, primarily causes disease on all susceptible apple cultivars (Jha *et al.*, 2010). Apple scab also known as black spot, is one of the most serious diseases of apple (Machardy, 1996) and could cause huge economic losses of up to 70% in the apple industry (Thakur *et al.*, 2013). The fungus also infects *Malus* (Crabapple), *Cotoneaster integerrima*, *Crataegus oxycantha* (Hawthorn), *Loquat*, *Pyracantha* (Firethorn), *Sarcocephalus esculantus*, *Sorbus* (Mountain Ash), and *Viburnum* (Jha *et al.*, 2010). Trees or orchards not treated with chemical sprays against apple scab have a sickly appearance with deformed leaves and fruits having irregular shape and size. Leaves and fruits also drop prematurely leaving the trees vulnerable to chilling and freezing injuries. Apple scab dates back as far as 1819 in a scientific report of Fries, nonetheless it was also illustrated in a sixteenth century

painting with apple scab infected apples to suggest its existence at that time (MacHardy *et al.*, 2001). In 1880, Winter classified the fungus into the *Venturia* genus.

1.3.1.1 Life cycle of *Venturia inaequalis*

Venturia inaequalis, a hemibiotrophic fungus, is a filamentous ascomycete which includes a parasitic phase as a subcuticular biotroph and a saprophytic sexual phase on infected fallen leaves (Machardy, 1996; Kucheryava *et al.*, 2008). Symptoms of the disease on apples include a circular olive green to velvety brown lesions, necrotic or chlorotic lesions single or scattered on leaf surface, olivaceous spots on infected sepals and pedicels (Fig. 1.2). On young fruits apple scab appears as corky lesions while on matured fruit as small black spots (Fig. 1.2).

(a)  (b)  (c) 

Figure 1.2 Symptoms of apple scab disease. The effect of *Venturia inaequalis* on (a) leaves and (b) young and (c) mature fruit (http://www.agf.gov.bc.ca/cropprot/tfipm/applescb.htm).

The life cycle discussed below is obtained from Bowen *et al.* (2011) and illustrated in Figure 1.3 in order to present a better understanding of the pathogen's propagation strategy. *Venturia inaequalis* produces sexual and asexual spores, which are capable of infection. The fungus survives the winter on dead leaves on the ground. During this overwintering time the fungus develops sexually which serve as the primary source of

inoculum. The mycelium of the fungus penetrates a bit deeper into leaf tissues and switches from vegetative to reproductive phase wherein the mycelia of two different mating types undergo sexual reproduction. Following mating, a pseudothecium is produced. The pseudothecium lumen starts to fill up with pseudoparophyses. Asci, which house ascopores, start to develop within the pseudothecium. For pseudothecia development to occur, moisture is needed. Ascopores consist of two unequal sized cells, hence the name of the fungus.

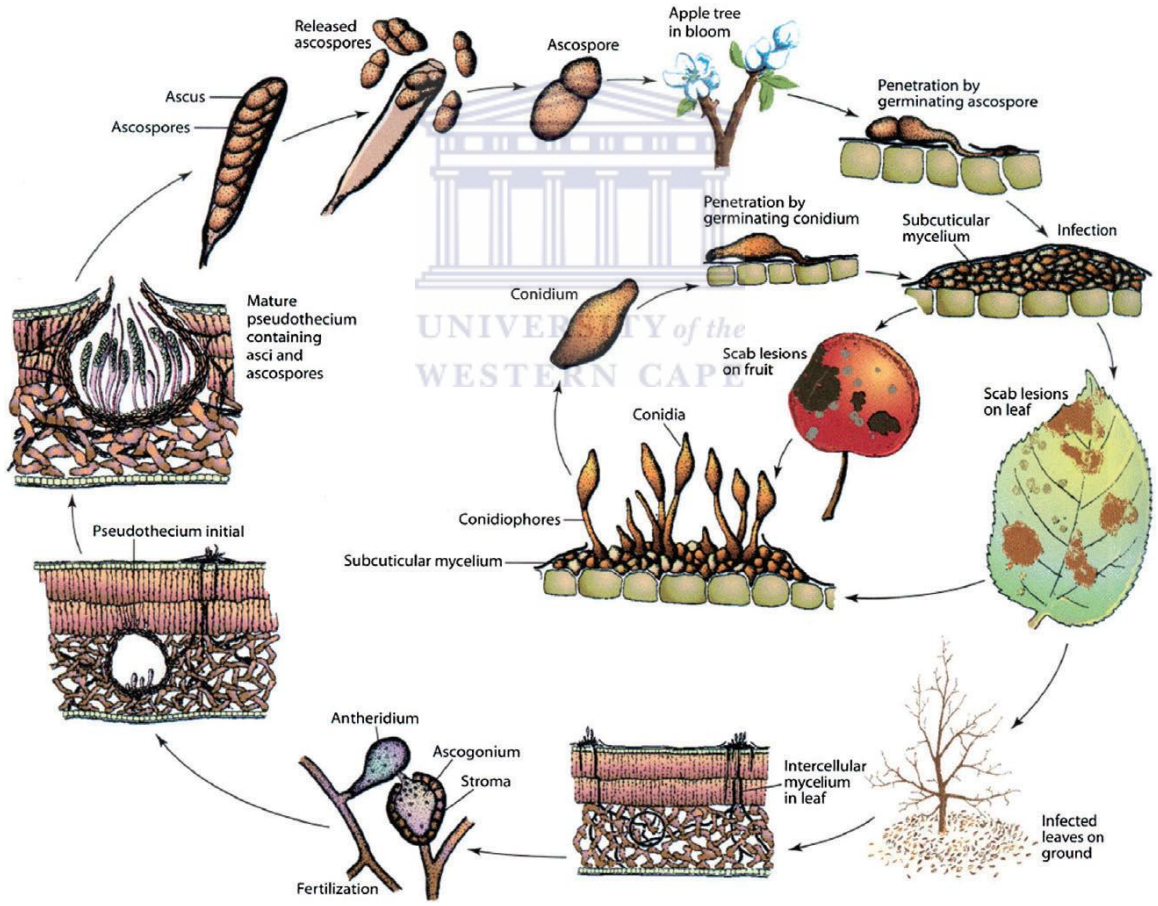


Figure 1.3 The life cycle of *Venturia inaequalis*. Figure illustrates the life cycle of the fungus, the asexual phase on young leaves and the saprophytic sexual phase on fallen leaves (Copyright John Wiley and Sons 2014).

The ascopores are released with the aid of splashing rain early in spring and dispersed by the wind to infect other young plants. The infection risk is higher early in the growing season when leaves and fruits are tender and most susceptible (Bowen *et al.*, 2011). Once it lands on the leaf or fruit, the ascospore germinates, in the presence of water, and infection occurs as hyphae penetrate the cuticles. A subcuticular stroma (Fig. 1.3) develops which is responsible to obtain nutrients from the subcuticular space, followed by the development of conidiophores (Bowen *et al.*, 2011). Conidia are formed on the conidiophores, with subsequent lesions being seen. The conidia are transferred to other parts of the tree by splashing rain and wind. The fungus subsequently infects the new leaf or fruit in a similar fashion to the ascopores. The rate at which events occurs, is temperature dependent. Happening less often in milder climates, conidia produced asexually, can overwinter in the dormant pustules on shoots and budscales are transferred to new plants where infection occurs. In this situation the primary inoculum are the conidia, but in most other cases the primary inoculum are the ascopores.

In order to infect the plant and propagate, the pathogen needs to adhere, germinate and form infection structures to penetrate the host (Kucheryava *et al.*, 2008). The conidia or ascospores germinate in the presence of moisture on the leaf by developing a germ tube. The fungus penetrates the waxy cuticle of the leaf but not through the stomata (Kucheryava *et al.*, 2008). Different hypotheses were investigated about the penetration of the fungus. Steiner and Oerke (2007) found that a melanized ring formed at the base of appressoria by the fungus was essential for infection. The possibility of applying mechanical pressure being used by the fungus was also excluded by (Smereka *et al.*,

1987). Furthermore, it was hypothesised that the fungus uses enzymatic hydrolysis to breach the cuticle. It is found that extracellular cutinases are produced by germinating conidia and mycelia (Koller *et al.*, 1991). Softening of the cuticle was ascribed to an esterase-like activity in the presence of germinating conidia for easy penetration of the fungus (Nicholson *et al.*, 1972). Upon penetration it establishes an infection hyphae and sub-cuticular stromata between the upper epidermis and the cuticular membrane without penetrating cells or invading intercellular spaces (Kucheryava *et al.*, 2008). The stroma gives rise to conidiophores, which bear conidia that bulge out from the host cells by rupturing the epidermis, hence the name apple scab.

1.3.1.2 Origin and spread of the apple scab disease

The reproductive structures of the fungus namely the ascospores and conidia (discussed in 1.3.1.1) have limited dispersal capacities with conidia being dispersed over a few meters and ascospores not more than a few kilometers. Therefore, long-distance dispersal of the disease would have been mediated through the transportation of infected fruits and plants (Gladieux *et al.*, 2008). Based on this premise, the pathogen will mirror the movements of its host. In a study done by Gladieux *et al.* (2008), they revealed that *Venturia inaequalis* emerged in Central Asia and have subsequently followed its host into Europe along the Silk Roads (Fig. 1.1). As apple production expanded to North and South America, Australasia and South Africa, the spread of *Venturia inaequalis* also increased and is now found in all apple growing regions (Gladieux *et al.*, 2008).

1.3.1.3 Disease management strategies

As mentioned in section 1.3.1.1, apple scab is the most serious disease of the apple industry and it is the cause of major economic losses. The primary inoculum of apple scab is from ascospores released from pseudothecia on overwintered infected leaves. Management of apple scab primarily relies on the use of fungicides between 15 and 25 applications annually depending on weather conditions, disease pressure and cultivar susceptibility (Holb *et al.* 2005b). It is also known that intensive fungicide spray programs can render some *Venturia inaequalis* strains resistant, which could become endemic. It has been reported that the fungus has successfully developed resistance to a number of fungicides, including dodine and benzimidazole, demethylation and quinone inhibitors (Koller, 1994). The need for much more environmentally friendly approaches towards apple scab management has become more critical. Amongst these, sanitation practices can contribute to resistance management since the fungus has become less susceptible to fungicides (Gomez *et al.*, 2007). Control of apple scab can be achieved through various approaches including sanitation, biological and chemical control strategies (Gomez *et al.*, 2007). Each of these approaches contributes some degree of success. However, when sanitation practices, biological- and chemical control, film forming polymers, resistant breeding and genetic modification strategies are combined, they have a better and longer effect on the management of apple scab.

1.3.1.3.1 Sanitation measures: leaf litter management

Ideally, eliminating the incidence of apple scab would primarily require the inhibition of pseudothecial development on overwintered foliage. Practices that indirectly reduce population density of *Venturia inaequalis* involve removal of leaf litter. Through this action the saprophytic sexual reproduction stage of the fungus that occurs on dead leaves is prevented and ultimately a reduction in ascospore discharged into the orchard air. Removal strategies could be the following: (i) physically removing fallen leaves from the orchard by hand or by machinery, (ii) raking and then plowing the leaf litter under the soil, (iii) raking and then burning the leaf litter, (iv) shredding the leaf litter with a flail mower, (v) treating the leaf litter with chemicals such as urea which suppress the sexual stage (Burchill *et al.*, 1965) and ultimately hasten decomposition (by making it more vulnerable to attack or by increasing the population density of decomposers), (vi) treating the leaf litter with microorganisms that decompose the leaves or grow on the leaf surface and physically prevent discharged ascospores from further flight into the orchard air, and (vii) utilizing earthworms to bury leaves. However, these practices can only reduce the primary inoculum and is usually not sufficient to control apple scab (Machardy, 1996).

1.3.1.3.2 Biological control

Bio-control may be defined as the use of insects or microbes to effectively control or inhibit the incidence of disease. The fungal antagonist, *Microsphaeropsis ochracea*, inhibits *Venturia inaequalis* growth through direct penetration and active growth within the hyphae, inducing fungal cell death (Benyagoub *et al.*, 1998). Vincent *et al.* (2004) found the most efficient treatment to reduce apple scab infection was to apply urea,

followed by leaf shredding, then applying *Microsphearopsis ochracea* and by *Athelia bombacina* Pers. inoculations. However, the use of urea and fungal antagonists are prohibited in Europe because it is not registered and therefore cannot be used in organic farming (Gomez *et al.*, 2007). Niklas and Kennel (1981) concluded that the earthworm, *Lumbricus terrestris* possesses the ability to consume the fruiting bodies of various fungi including that of *Venturia inaequalis*. In quite a number of studies it was shown that *L. terrestris* plays a key role in apple leaf litter decomposition. Furthermore it was also observed that softening of leaf material using urea and other nitrogeous compounds led to an increase of earth worm activity (Holb *et al.* 2006). It was also observed that some fungicides such as benzimidazoles (benomyl and thiophanate-methyl), dodine and captafol applied to fallen leaves had a toxic effect on earthworm populations. On the other hand captan sprays had no effect on earthworm populations. Thus, earthworms as such may be used as an effective control mechanism for apple scab provided that it is used in conjunction with bio-friendly fungicides.

1.3.1.3.3 Chemical control

Chemical control is fore mostly attained through the use of fungicides. Fungicides are divided into two classes: protectant fungicides and post-infection fungicides. The former prevents spore germination and inhibits cuticle penetration while the latter regulates fungal growth within host tissues, confining it to the localized area of infection. A minimum of 15 applications (Table 1.1) is required per growing season at vital developmental stages. Fungicidal control of apple scab may be effective if application is perfectly timed and regular. This approach has proven to be successful on susceptible cultivars. Timing of the first application is critical for early infection of apple scab. This

is very important for control of later leaf and fruit infections. Fungicide applications can be made as early as the green tip growth stage (Table 1.1) and continued until petal drop.

The amount of treatments required, deviates when weather conditions suggest that a scab out break may be imminent, prompting its immediate application. In extreme cases apple crops are sprayed up to 18 times with a combination of fungicides (Holb *et al.* 2005). A wide spectrum of fungicides are commercially available, a selection of which is used at the farmers' discretion. Factors that influence the selection process include the orchards disease history, socio-economic aspects affiliated with the farmer and the desired final market (Biggs, 1997).

Table 1.1 Apple tree growth stage

Growth stage	*Description*
Dormant	Absence of growth in autumn, winter or spring
Silver Tip	Swollen buds become noticeable and the silvery fuzzy leaf tissue begins to emerge from the tip of the bud.
Green Tip	Green leaf tissue is visible at the tip of the bud.
Tight Cluster	The spur leaves have folded back exposing the flower cluster inside the bud.
Pink/Pre-bloom	The flower buds have grown enough to expose the petals of the flowers.
King Bloom	The center "King Bloom" has opened.
Full Bloom	Eighty percent or more of the flowers on the tree are open.
Petal Drop/Fall	Last petals have fallen from blossoms.

1.3.1.3.4 Film forming polymers an alternative management disease strategy

An alternative to the disease management strategies discussed in 1.3.1.3.1, 1.3.1.3.2, and 1.3.1.3.3, namely the application of film forming polymers, gained new attention. This technology has been successfully applied and is widely used as spray adjuvants within the agricultural, forestry and horticultural industries (Backman, 1978). It was initially developed to (i) reduce weathering and extend pesticide efficacy, (ii) to act as stickers/spreaders to improve distribution and adherence of agrochemicals, and (iii) to decrease water loss and wilting of young transplants (Gale and Hagan, 1996). The use of forming polymers proofed to be more effective in the control of leaf pathogens of various crops *viz.* cereals, vegetables, fruit and ornamentals, including rusts (*Puccinia*, *Uromyces* spp.), grey mould (*Botrytis cinerea* Pers.), eyespot (*Septoria nodorum* Berk.) and leaf spot (*Pyrenophora* spp.) (Sutherland and Walters, 2001; Sutherland and Walters 2002; Blaedow *et al.*, 2006). These polymers act as a physical protective barrier against invading of pathogens at the leaf surface, prevent adherence of spores and subsequent inhibition of germ tube development (Osswald *et al.*, 1984; Han, 1990). The advantage that film forming polymers has, it act by physical means and therefore not subjected to the stringent government legislative restrictions that relate to the use and application of conventional pesticides (Percival and Boyle, 2009). Film forming polymers are less phytotoxic than synthetic fungicides to leaf tissue, permeable to atmospheric gases and allow penetration of solar radiation with little detrimental effects on tree biology (Fuller *et al.*, 2003). These polymers offer a cheaper alternative at a cost of 40-80% less than conventional fungicides and are an environmentally more acceptable system for reducing

scab related yield losses. Percival and Boyle (2009) conducted an apple scab infection study and found for both *in vitro* and field trials that the film forming polymers was quite successful in the reduction of apple scab severity. Their results also showed an increase in apple yield per tree for both years of the trial. This technology can thus be regarded as a safe alternative to the already existing methods of apple scab management (Percival and Boyle, 2009).

1.3.1.3.5 Breeding for apple scab resistance cultivars

The high number of chemical sprays applied on apple crops raised concerns with consumers about the environment as well as food safety (Broggini *et al.*, 2010). This awareness is reflected in legislation followed by the continued pressure from retailers to reduce residue levels on fresh produce (Broggini *et al.*, 2010). These circumstances constrained plant breeders to breed new durable apple scab resistant cultivars with a range of resistance genes at their disposal (Gessler and Patocchi, 2006). Breeding of apple scab varieties through genetic resistance is the preferred method for controlling disease. Research groups aim to introduce into the host resistance genes to various diseases. This was challenged by the number of diseases, which made it difficult to obtain resistance to all diseases in one cultivar. However, the different resistant varieties exhibiting monogenic resistance (Table 1.2) was also overcome by apple scab hence gene pyramiding was proposed. Gene pyramiding is the combination of different resistant genes in the same genotype exhibiting durable resistance. Breeding programs started as early as 1944 by Hough with the discovery of the apple scab resistant clone of *Malus floribunda*. This resistance was attributed to the dominant gene *Rvi6* (*Vf*) (Hough *et al.*,

1953). As a result of this discovery, breeding programs were intensified resulting in over 70 new varieties being produced (Crosby *et al.*, 1992). A number of other resistance genes against apple scab have been identified and several sources of polygenic resistance are known (Table 1.2).

Table 1.2 A reviewed list of apple scab resistance genes and subsequent plant sources and matching avirulence genes.

Malus					***Venturia inaequalis***				
Representative host		**Phenotype**	**Resistance locus**		**Avirulence locus**		**Representative isolate**		
No.	Accession		Old	New	New	Old	Name	Accession	Reference
h(0)	(Royal) Gala	Susceptibility					(0)		
h(1)	Golden Delicious	Necrosis	*Vg*	*Rvi1*	*AvrRvi1*		(1)	EU-B04	(Benaouf, Parisi, 2000)
h(2)	TSR34T15	Stellate necrosis	*Vh2=Vr*	*Rvi2*	*AvrRvi2*	p-9	(2)	1639	(Bus *et al.* 2005)
h(3)	(F1 of Geneva)		*Vh3*	*Rvi3*	*AvrRvi3*	p-10	(3)	1774-1	Bus, unpublished data
h(4)	TSR33T239	Hypersensitive response	*Vh4=Vx=Vr1*	*Rvi4*	*AvrRvi4*		(4)		
h(5)	9-AR2T196	Hypersensitive response	*Vm*	*Rvi5*	*AvrRvi5*		(5)	163	Hernandez *et al.*, 1994
h(6)	Priscilla	Chlorosis	*Vf*	*Rvi6*	*AvrRvi6*		(6)	302	(Benaouf, Parisi, 2000)
h(7)	F1 of *M. x floribunda*	Hypersensitive response	*Vfh*	*Rvi7*	*AvrRvi7*		(7)	1066	(Benaouf and Parisi, 2000)
h(8)	GMAL3631-W193B	Stellate necrosis	*Vh8*	*Rvi8*	*AvrRvi8*		(8)	NZ188B.2	(Bus *et al.* 2005)
h(9)	J34 (F1 of Dolgo)	Stellate necrosis		*Rvi9*	*AvrRvi9*		(9)	1639	Bus, unpublished data
h(10)	Antonovka PI172623	Hypersensitive response	*Va*	*Rvi10*	*AvrRvi10*		(10)		
h(11)	*M. baccata jackii*	Stellate necrosis/chlorosis	*Vbj*	*Rvi11*	*AvrRvi11*		(11)		
h(12)	Hansen's baccata	Chlorosis	*Vb*	*Rvi12*	*AvrRvi12*		(12)		
h(13)	(F1 of) Durello di Forli	Stellate necrosis	*Vd*	*Rvi13*	*AvrRvi13*		(13)	EU-NL05	(Parisi *et al.*, 2004)

1.3.1.3.6 Application of molecular biological techniques

The first line of thought to increase resistance in apple scab-susceptible commercial cultivars would be to genetically modify these cultivars with genes from apple scab resistant cultivars. Plant breeders have been doing cross breeding of susceptible varieties with resistant varieties over the past 70 years since the discovery of the *Rvi6* (*Vf*) gene in the *Malus floribunda* clone (section 1.3.1.3.5). They also bred for improved fruit texture such as crispness, firmness and juiciness (Mckay *et al.*, 2011). However through these conventional methods it often took years before the desired effect or trait could be observed. This process can be expedited by adding resistance to an already existing high quality cultivar through marker-assisted "fast" breeding (Volz *et al.* 2000) or cis-genesis (Schouten *et al.*, 2006). Molecular tools made it possible to introduce a gene of interest directly into the crop.

Most of the conventionally bred cultivars could resist races 1 to 5 (Szankowski *et al.*, 2009) and race 8. The first report of apple scab to overcome *Rvi6* (*Vf*) resistance was in the Prima cultivar in Ahrensburg, Germany, by Parisi *et al.* (1993), which they referred to as race 6 (Parisi *et al.*, 1993). They also discovered another isolate which overcomes *Rvi6* (*Vf*) resistance and called it race 7 (Parisi *et al.*, 2004). With the breaking down of *Rvi6* (*Vf*) resistance obtained through classical breeding, genetic transformation of apple varieties was necessitated.

The identification and mapping of several apple scab resistant genes such as *HcrVf* (Homologues of *Cladosporium fulvum* resitance genes of *Rvi6* (*Vf*) region) (Patocchi *et*

al., 1999), *Rvi15* (*Vr2*) (Patocchi *et al.*, 2004), *Rvi13* (*Vd*) (Soriano *et al.*, 2009), *Rvi11* (*Vbj*) (Gygax *et al.*, 2004) on different linkage groups of the apple genome, made genetic transformation of scab susceptible varieties possible. Xu and Korban (2002) showed that the *Rvi6* (*Vf*) region consists of a cluster of four putative resistant genes, named *HcrVf1*, *HcrVf2*, *HcrVf3* and *HcrVf4* (Xu and Korban, 2002). With this information in hand, Belfanti *et al.* (2004) genetically transformed the susceptible Gala cv. with the *HcrVf2* gene, which they assumed by analogy with the *Cf* gene in tomato to serve in pathogen recognition according to the gene-for-gene rule. The resulting transgenic apple plants, in which the gene was over-expressed, proved to be resistant to apple scab.

More and more research groups do genetic transformations on susceptible cultivars for example transforming 'Galaxy' and 'McIntosh' with *Vfa1* and *Vfa2* using native promoters and terminators (Malnoy *et al.*, 2008). More recently, Joshi and Schaart (2011) reported on the role and resistance spectrum of the *cis* genes *HcrVf1* and *HcrVf2* also using native promoters and terminators. Malnoy *et al.* (2008) showed successfully partial resistance against a mixture of isolates of *Venturia inaequalis* and transformants generated by Joshi and Schaart (2011) reached scab resistance levels comparable to the *Vf* resistant cultivar Santana, which were obtained through classical breeding.

An alternative to *trans* or *cis*-genic obtained resistance is resistance acquired through systemic resistance. Increasing disease resistance through the expression of genes which have a cascading effect on pathogenesis-related (PR) proteins leading to systemic acquired response (SAR), has been obtained in other plant crops such as *Arabidopsis*

thaliana, rice, tomato and wheat and apple (Malnoy *et al.*, 2007). In this study by Malnoy *et al.* (2007), they showed that disease resistance could be acquired through genetic engineering of plant defence-related genes or pathogen-originated genes. They have targeted SAR, which is long lasting and effective against a broad spectrum of pathogens.

1.4 Different scab races on different apple cultivars

The term isolate or physiological race of a *Venturia inaequalis* isolate has been debated (Gessler *et al.*, 2006). They have come to the conclusion that "a race designates an isolate capable of infecting and finally sporulating on a particular host resistant to other isolates" and should therefore be termed a physiological race (Gessler *et al.*, 2006). This determination of the race is based on the theory that a mutation occurred at the *Avr* locus of the pathogen leading to the non-recognition of the host; hence complete susceptibility (Bus *et al.* 2011). In concordance with Gessler's definition, eight different physiological races have been identified (Machardy, 1996; Benaouf and Parisi, 2000; Bus *et al.*, 2005) for the pathogen based on their hypervariability and the exhibition of differential pathogenicity on different apple cultivars. The most important features of these races to date in literature have been summarized in Table 1.3. It was also observed that some of *Venturia* isolates infect two different differential hosts, which makes the classification of the particular race difficult. Bus *et al.* 2009 proposed a new nomenclature system by renaming the resistance locus in the different hosts, thus the old way of naming a particular host for eg. *Vf*6 of *Malus floribunda* 821 has been changed to *Rvi*6, making the system more robust.

Table 1.3 Apple scab races affecting different apple cultivars

Races	*Pathological characteristic on apple cultivars*	*References*
Race 1	Non sporulating lesion on Dolgo, R 12740-7A (a Russian cultivar) and Geneva and most of the cultivars worldwide	(Machardy, 1996)
Race 2	Sporulating lesions on Dolgo, Geneva and some progenies of R 12740-7A	(Machardy, 1996)
Race 3	Sporulating lesions on Geneva, and non sporulating lesion on Dolgo, R 12740-7A	(Machardy, 1996)
Race 4	Non sporulating lesion on Dolgo, Geneva and sporulating lesion on those progenies of R12740-7A on which race 2 isolates cannot sporulate	(Machardy, 1996)
Race 5	Sporulating lesions on *Rvi5* (*Vm*) R gene containing cultivars	(Machardy, 1996)
Race 6	Sporulating lesions on *Rvi6* (*Vf*) hybrids but cannot infect *Malus floribunda* 821 containing *Rvi7* (*Vfh*) R gene	(Benaouf and Parisi, 2000)
Race 7	Can infect cultivars having *Rvi6* (*Vf*) and *Rvi7* (*Vfh*) R gene but cannot infect Golden delicious which contains *Rvi1* (*Vg*) gene	(Benaouf and Parisi, 2000)
Race 8	Can infect Golden delicious, Royal gala, and cultivars containing *Rvi8* (*Vh8*) R gene	(Bus *et al.*, 2005)

1.5 Types of Resistance in apple

Apple scab causes severe disease symptoms on apple as discussed in section 1.3.1.1. Plants, in particular apple, are constantly under attack by a variety of microbes and fungi including *Venturia inaequalis*. Apples evolved strategies to recognise these pathogens with defence mechanisms being activated (Cova *et al.*, 2009). This could be attributed to protein-protein interactions which involve pathogen-encoded elicitors namely avirulence (Avr) proteins and plant host receptors which are encoded by resistance (R) genes. This association is known as the gene-for-gene mode of interaction in which major genes in the host interact with corresponding avirulence genes (Table 1.2) in the pathogen (Benaouf and Parisi, 2000). This is based on the hypothesis of Flor (1955), which states that for every gene in the plant that confers resistance, there is a corresponding gene in the pathogen that confers avirulence. This further implies that, after pathogen recognition, a cascade of cytoplasmic and nuclear responses is induced that ultimately leads to the regulation of various defence-related genes. These genes bring about the following: depositing of callose rich appositions or papillae at sites of attempted pathogen penetration, accumulation of phenolic compounds and various toxins in the cell wall and finally synthesis of lignin-like polymers to reinforce the wall (Cherk *et al.*, 2009).

Resistance to the pathogen can be polygenic with several genes involved or monogenic where only one single gene is involved. Types of resistance include ontogenic and race specific or race non-specific resistance. Ontogenic resistance refers to the intensity of the symptoms, which is a progressive increase in incubation time up to no symptom appearance (Gessler *et al.*, 2006). This resistance runs concurrent with leaf aging, which

result in a suppression of fungus growth. However, it is speculated that the strengthened cell wall, cuticular membrane along with sub-cuticular pH of these leaves play a role in restricting fungal growth. A study done by Gusberti *et al.* (2013) concluded that the modulation of five candidate genes contributed to ontogenic resistance in apple.

For race specific or non-specific resistance, Machardy (1996), Gessler (2006) and Chevalier (1991) observed that the isolates or races of *Venturia inaequalis* induce variable symptoms on different apple cultivars. These responses were classified according to Hough *et al.* (1953) who devised a quantitative 5-class system for scab resistance. Chevalier *et al.* (1991) redefined the classification model and is still the preferred model today. Classes 0 and M used in Hough's system were done away with and class 3 was split into 3a and 3b. Their classification consisted of class 1, 2, 3a, 3b and 4. They observed this phenomenon in the progeny of an *Rvi6* (*Vf*) selection crossed with a commercial (susceptible) cultivar. He however, characterised his class 1 by the typical hypersensitive or pin-point response, with extended and rapid cell death, including changes in the palisade tissue in the cells in contact with the pathogen. Furthermore, class 2 had chlorotic lesions with a slightly necrotic centre, class 3a consisted of necrotic and

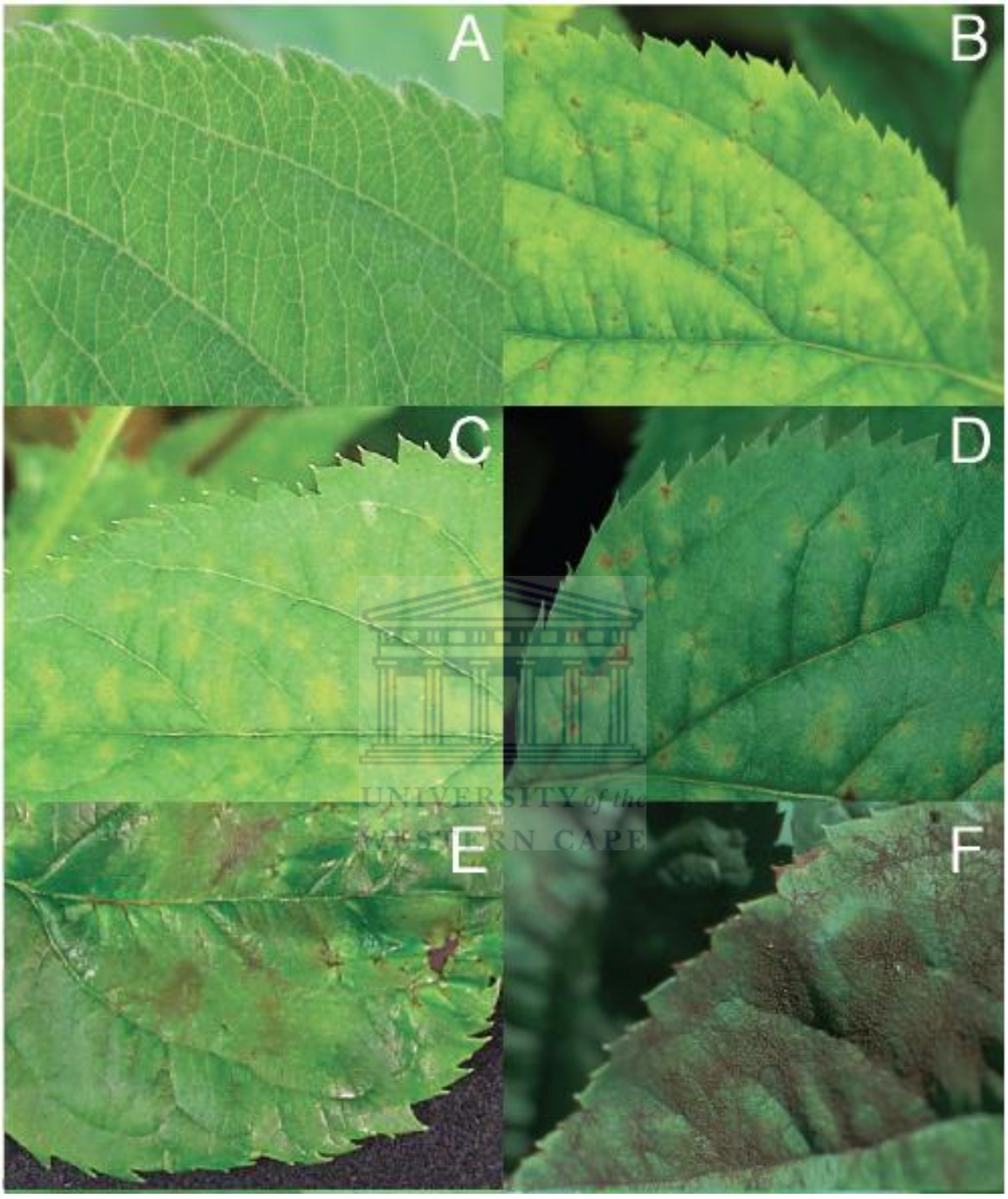


Figure 1.4 Representation of scab reaction classes after *Venturia inaequalis* inoculation (about 21 days post-inoculation): A. class 0 or no symptoms; B. class 1 or 'pinpoint pits' (e.g. *Vm and Vh4*); C. class 2; D. class 3a; E. class 3b; F. class 4 or complete susceptibility (Gessler and Patocchi, 2006, Copyright Taylor and Francis Group 2014).

some chlorotic lesions with occasional very slight sporulation, class 3b had very clear sporulating chlorotic and necrotic lesions. Class 4 was characteristic of abundantly sporulating lesions. Simply stated, class 1 is most resistant, class 2 is resistant, class 3a is weakly resistant, class 3b is weakly susceptible and class 4 is the most susceptible (Fig. 1.4).

1.6 Plant defence mechanisms

Plants are constantly under attack from pathogens and as a result developed a defence system in place. These microorganisms trigger activation of this system. They recognise these microbial pathogens through pathogen-associated molecular patterns (PAMPS) by sensors, which are known as pattern-recognition receptors (PRRs) that trigger a series of defence responses known as PAMP-triggered immunity (PTI; Chen, 2011). The mitogen-activated protein kinase (MAPK) cascades play important roles in mediating PAMP signals. Kishi-Kaboshi *et al.* (2010) have shown that MAPK cascade were triggered in response to a rice fungal infection. This is followed by the activation of transcription factors such as the MYB-family and WRKY pronounced "worky" transcription factors. In a study done by a Swiss group where peach was inoculated with *Xanthomonas arboricola* pv. *pruni*, transcription factors were up-regulated (Socquet-Juglard *et al.*, 2013).

The host also invest in reinforcing its physiological barriers such as the cell wall structure. This is achieved by inducing genes involved in cell wall modification such as invertase/pectin methylesterase inhibitor (Socquet-Juglard *et al.*, 2013), beta-glucosidases (Escamilla-Trevino *et al.*, 2006), peroxidase and lignin synthesis genes such as

phenylalanine ammonia lyase (PAL) and pectin peroxidase (Gusberti *et al*., 2013). The outcome of this combined effort causes starvation of the fungus and subsequently stops the spreading of the disease. Furthermore, genes encoding hydrolysis enzymes such as chitinase are also induced in order to hydrolyse the fungal cell wall. The presence of these proteins indicates that major changes are performed at an early stage by either degrading or lignifying cell walls in order to prevent the spread of the fungus to the rest of the host.

In most cases, a global down regulation of the photosynthetic genes (Bilgin *et al.*, 2010), a strategy adopted by the plant to limit availability of sugars for the pathogen and all the energy is reallocated for defence. It was hypothesised that the down regulation of photosynthesis genes could lead to the induction of a hypersensitive response blocking-off further spread of the fungus. Genes involved in hormone signalling pathways are also induced such as those in salicylic acid (SA), jasmonates (JA) and ethylene. This defence related signalling molecules play an active role in plant defence.

Microorganisms capable of causing disease, secrete effectors into the host, which interfere with the host cell function. The effectors are recognised in the plant by resistance (R) genes, resulting in extracellular oxidative burst, induction of SA signalling pathway, calcium and hydrogen influxes in the cell, cell death or hypersensitive response. R genes, which were induced or repressed include receptor-like kinases (RLKs), leucine-rich repeat (LRR) kinases. The presence of these proteins (RLKs and LRRKs), were reported in the study of Socquet-Juglard *et al.* (2013). Defence-related genes encoding proteins such as thaumatin, glutathione S-transferases, which serve as antioxidants for

cleaning the cell of ROS are also induced. These serve in protecting the plant from oxidative damage.

Genes encoding chemical defences such as pathogenesis related (PR) proteins are also induced in order to defend the plant against microbial attack. From this we can conclude that the plants have quite an arsenal to combat any bacterial, fungal or viral invasion but yet we get fungi, which overcome these.

1.7 Project approach

The advent of high-speed sequencing has opened a new door for the 'omics'. Proteomics and transcriptomics as tools to investigate host-pathogen interaction in animal and plant organisms have been extensively applied. In transcriptome studies, utilisation of next generation sequencing has opened the door for a global analysis of the mRNA complement of any cell. Changes in mRNA abundance are not always mirrored by the corresponding protein levels and early rapid changes in cell behaviour often rely on pre-existing proteins that are post-translationally modified or have changed their location and/or are degraded (Quirino *et al.*, 2010). In addition to this, more than one protein can be produced as a result of genes encoding differently spliced mRNA's. Since the presence of a gene or its mRNA is no guarantee of a role in cellular activity, therefore approaching this study at a protein level as well, is a welcoming complement to the transcriptomic element of this study. Shortcomings as listed for transcriptomics exist when utilised in isolation, so to approach this project using both tools will make it possible to cover almost every aspect.

1.7.1 Proteomics Approach

1.7.1.1 What is proteomics?

Two-dimensional electrophoresis (2-DE) was developed more than two decades before the term proteomics was coined (O'Farrel, 1975). Proteomics refers to the study of the proteome, which is defined as the global set of cellular proteins expressed in a particular biological state of the organism (Kim *et al.* 2007.). This method remains the method of choice for its unique ability to separate hundreds or thousands of proteins at a time. The 2-DE entails the separation of complex protein mixtures in two steps *viz.* (i) according to molecular charge in the first dimension and (ii) by molecular weight (MW) in the second dimension. This tool provides a systematic understanding of events at the molecular level according to Kim *et al.* (2007). It can be used for protein profiling, comparative expression analysis of two or more protein samples, the localisation and identification of posttranslational modifications and for the study of protein-protein interactions. For this study, the focus will be on comparative expression analysis between two samples. Two types of techniques *viz.* gel-based and gel-free techniques can be utilised.

1.7.1.2 Gel-based and gel-free proteomic techniques

A number of proteomic approaches have been developed over the past thirty years since the first report of protein separation in 1975 (O'Farrel, 1975). These approaches include gel-based applications that include one-dimensional and two-dimensional polyacrylamide gel electrophoresis and gel-free high throughput screening technologies. Gel-free

screening technologies are used to perform relative quantification of proteins by isotope labelling.

1.7.1.3 Proteomics workflow in plants: a gel-based approach

In order to use proteomics as a tool, proteins have to be isolated from the organism or tissue in question. This is then followed by a gel-based approach, which include the following: protein solubilisation, first dimension which is isoelectric-focusing (IEF) with immobilized pH gradients (IPGs), the second dimension (SDS-PAGE), gel staining, quantitative analysis of 2-DE images/or protein profiling and protein identification.

1.7.1.3.1 Protein isolation

Firstly it involves the extraction of the total protein from leaves, stems or roots, which is a challenging exercise. This is due to the presence of cell wall polysaccharides and polyphenols but also proteases, which can degrade samples. Protein extraction from plant material often requires physical disruption of the cell wall by mechanical means (grinding in liquid nitrogen) or sonication. Precipitation of proteins includes a trichloroacetic acid (TCA) and/or acetone protocol (Jacobs *et al.*, 2001) or phenol based extraction where proteins are solubilised in the phenolic phase and precipitated with methanol and ammonium sulphate (Carpentier *et al.*, 2005; Quirino *et al.*, 2010).

1.7.1.3.2 Solubilisation and separation

Second dimension electrophoresis of proteins have a high resolution capacity, therefore proteins have to be completely denatured, disaggregated, reduced and solubilised to disrupt molecular interactions to ensure that each spot represents an individual polypeptide. This is achieved by using a buffer containing chaotropes (urea and thiourea), zwitterionic detergents (CHAPS), reducing agents (DTT), carrier ampholytes and protease and phosphatase inhibitors (Rabilloud and Lelong, 2011). Total soluble protein (TSP) are separated via one- (1-DE) and second dimension (2-DE) electrophoresis, a technique developed by O'Farrel (O'Farrel, 1975). In 2-DE PAGE, protein sample are firstly separated via the isoelectric focusing point (based on protein p*I*) on an immobilized pH gradient (IPG) strip that has a gradient of pHs in the first dimension. This is followed by separation by mass on a SDS-polyacrylamide gel electrophoresis (PAGE) in the second dimension (Quirino *et al.*, 2010).

1.7.1.3.3 First dimension separation: IEF with IPGs

Proteins are amphoteric molecules, i.e. they carry positive, negative or zero net charge. The net charge of a protein refers to the sum of all the negative and positive charges. The isoelectric point (p*I*) of a protein is the specific pH at which the net charge of the protein is zero. The IPG strip allows separation of proteins in the first dimension based on their isoelectric point (p*I*). Prior to use, the IPG strip is rehydrated in a buffer containing carrier ampholytes and total soluble protein in solubilisation buffer. An electric field is applied in order to focus the proteins i.e. the negatively charged proteins move towards

the anode, the positively charged proteins move towards the cathode. Proteins are considered focused when they align according their p*I*, the global net charge is zero and unable to move. IEF is followed by equilibrating the proteins and this is achieved in two steps: (i) with an equilibration solution containing DTT to maintain a reducing environment and (ii) an equilibration solution containing iodoacetamide to alkylate reduced thiol groups, preventing their re-oxidation during electrophoresis.

1.7.1.3.4 Second dimension: SDS-PAGE

In the 2-DE SDS-PAGE, migration of proteins is determined by their molecular weight. The SDS-denatured and reduced proteins are separated according to molecular weight in comparison with a molecular weight marker. 2-DE analysis provides several types of information about the hundreds of proteins investigated simultaneously, including molecular weight, p*I* and quantity, as well as possible posttranslational modifications. 2-DE is extensively used but mostly for qualitative experiments. This method falls short in its reproducibility, inability to detect low abundant and hydrophobic proteins, low sensitivity in identifying proteins with pH values too low (pH < 3) or too high (pH > 10) and molecular masses too small (Mr < 10 kD) or too large (Mr > 150 kD).

1.7.1.3.5 Staining methods

In order to visualise the resolved proteins, gels must be stained. Firstly, the gel must be immersed in a fixation solution containing acetic acid and ethanol/methanol. Proteins are than visualised by staining gels with Coomassie Brilliant Blue (Neuhoff *et al.*, 1988),

silver nitrate (Chevallet *et al.*, 2006), or a number of fluorescent dyes (Berggren *et al.*, 2000; Luche *et al.*, 2007). Silver nitrate has a greater sensitivity but it does not show linearity of signal and is less compatible with MS and therefore not recommended for a gel comparison study. A formaldehyde-free silver staining method such as Pro-Q Diamond and Pro-Q Emerald for detection of phosphoproteins and glycoproteins, which is compatible with MS has been developed (Zhang *et al.*, 2013). Fluorescent dyes such as Sypro Ruby, Flamingo Pink™ and Deep Purple are very sensitive and compatible with MS technology and are commercially available. Sypro Ruby allows for the detection of lipoproteins, glycoproteins, metalloproteins, calcium-binding proteins, fibrillar proteins, and low molecular weight proteins.

1.7.1.3.6 Proteome profiling and identification

Following staining of gels, whether it's conventional or fluorescent dyes, the gels are analysed and compared. The image of the gel is imported into specific software for comparative analysis. In order to perform a comparative analysis, at least three repetitions of the same sample should be run to reduce gel variability. Several commercial software packages are available including Image Master (Nonlinear Dynamics, Newcastle upon Tyne, UK), Progenesis Same Spots (Nonlinear Dynamics), PDQuest (Bio-Rad Laboratories, Hercules, CA, USA) and Melanie (Geneva BioInformatics, Geneva, Switzerland) can be used to perform a comparative analysis.

Protein spots deemed differentially expressed, based on statistical analysis of the gels, are excised and processed for identification by MS analysis. The processing of the proteins

involves an “in-gel-digestion” in which the protein, still in the matrix, is digested with trypsin that cleaves at specific points. Trypsin cleaves the peptide chain at the carboxyl side of lysine and arginine before proline. The peptide products are then introduced into the mass analyser. The proteins are identified through peptide mass fingerprinting (PMF) or via tandem MS analysis by *de novo* sequencing. In PMF, absolute masses of the peptides obtained from the mass spectrometer (Clauser *et al.*, 1999), are then compared by bioinformatics to a database containing known protein sequences or the genome of the organism. Computer programs such as MASCOT are designed to translate genomes into proteins, theoretically cleave the proteins into peptides. The absolute masses of the peptides from each protein are calculated automatically. The peptide masses of the unknown target protein are then compared to the theoretical peptide masses deposited in the database.

Tandem mass spectrometry (MS/MS) or *de novo* sequencing involves the selection of an ion using a mass filter or analyser. This selected ion is then fragmented further and analysed. The fragmentation is done in an argon gas filled chamber where the collision of the argon atoms and the ions result in fragmentation. The resulting daughter ion spectrum is then analysed. The protein sequence/s obtained can now be submitted to a database for identification. This approach is often the alternative to PMF where the protein sequences are not present in the database utilised (Quirino *et al*., 2010).

1.7.1.4 Protein analysis

Proteins and peptides are polar, non-volatile and thermally unstable and as a result of their nature, it require an ionization technique in which the analyte is transferred into a gaseous phase without being extensively degraded (Yates *et al.*, 2009). Matrix-assisted laser desorption ionization (MALDI) (Karas and Hillenkamp, 1988) and electrospray ionization (ESI) (Fenn *et al.*, 1989) are soft ionization methods that allow proteins to be analysed by mass spectrometry (MS). Briefly, the MALDI matrix absorbs laser energy and this is transferred to the acidified analyte, whereas the rapid laser heating causing desorption of the matrix and $[M+H]^+$ ions of analyte into the gas phase. This ionization requires several hundred laser shots to achieve an acceptable signal-to-noise ratio for ion detection. Ions generated through MALDI are singly charged, which makes it applicable to top-down analysis of high-molecular-weight proteins with pulsed analysis instruments (Yates *et al.*, 2009).

ESI is different from MALDI in that it produces ions from a solution. It is driven by high voltage (2-6 kV) applied between the emitter at the end of the separation pipeline and the inlet of the mass spectrometer. Physicochemical processes of ESI involve the creation of electrically charged spray, Taylor cone, followed by the formation and desolvation of the analyte-solvent droplets (Yates *et al.*, 2009). The latter is aided by a heated capillary or by sheath gas flow at the mass spectrometer inlet. The method is very sensitive as a result of the lowered flow rates from microliter-per-minute to nanoliter-per-minute. This source is usually coupled to the continuous analysis instrument.

1.7.1.5 Instruments utilised in proteomics

There are quite a number of instruments on the market that can be utilised to analyse proteins. There are two categories *viz.* (i) scanning mass spectrometers such as the TOF and quadrupole (Q) and (ii) the ion-beam mass spectrometers such as the ion trap (IT), Orbitrap and FT-ICR. The scanning mass analysers like the TOF are interfaced with the MALDI to perform pulsed analysis. The ion beam and trapping instruments are coupled to a continuous ESI source. The following instrument configurations are the most widely used solutions in the field of proteomics: ion traps [QIT: three-dimensional (3D) ion trap, LIT: linear ion trap], triple quadrupoles (TQ), LTQ-Orbitrap hybrid instrument (Thermo Scientific), LTQFTICR (Thermo Scientific), and the TQ-FTICR hybrid instruments Q-TOF and IT-TOF (Shimadzu; Yates *et al.*, 2009). The LTQ-Orbitrap hybrid instrument was used in this study and discussed in detail in the review of Yates *et al.* (2009).

1.7.1.6 LTQ-Orbitrap hybrid instrument

The LTQ-Orbitrap Velos^{TM} uses a novel mass analyser. It uses orbital trapping of ions in its static electrostatic fields in which the ions orbit around a central electrode and oscillate in an axial direction (Fig. 1.5).

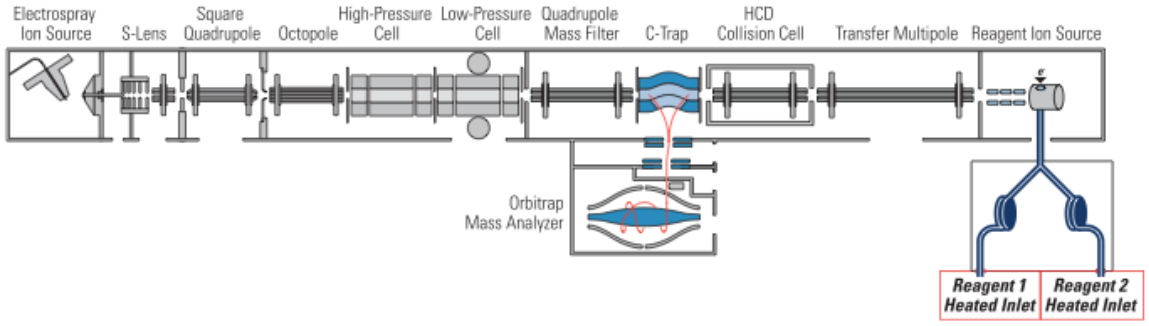


Figure 1.5 Schematic diagram of the LTQ Orbitrap Velos™ (Yates *et al.*, 2009).

The Orbitrap mass analyser has become quite popular for many proteomic applications since its launch in 2005. The ability to deliver low-ppm mass accuracy and extremely high resolution makes the Orbitrap the instrument of choice in many proteomic applications. Since 2005, several papers have been published in which the Orbitrap was successfully utilised in proteomic applications covering from human proteomics (Schenk *et al.*, 2008) through to animals (Mann and Mann, 2011), mouse liver and brain (Wisniewski *et al.*, 2009), pathogens (Valcu *et al.*, 2009) and plants (Carli *et al.*, 2011).

1.7.1.6.1 Gel-free proteomics

The following techniques are referred to as gel-free or MS based proteomics techniques. These include isotope-coded affinity tag (ICAT), isotope-coded protein label (ICPL), stable isotope labelling with amino acids in cell culture (SILAC), stable isotope labelling with amino acids in mammals (SILAM), isobaric tag for relative and absolute quantitation (iTRAQ), ^{18}O stable isotope labelling, dimethyl and tandem mass tags (TMT). These techniques have become the methods of choice for quantitatively comparing protein levels among biological proteomes. They are more sensitive and

reproducible than 2-DE gel-based methods. However, this section of the literature review will be limited to gel-based proteomics and its application in plants.

1.7.1.7 Application of proteomics in plants

With more and more genomes being sequenced, proteomics as a tool has become popular in unraveling biochemical and physiological mechanisms of complex multivariate diseases at the functional molecular level in all living organisms. Proteomics are used in a diverse field of biology including medicine, drug discovery, agriculture and to ecology and population biology. Biotic such as pathogens (Savory *et al.*, 2012) and abiotic (drought and salt) stresses and the effects these stresses have on plants or animals on a molecular level are being elucidated. Furthermore, proteomics are being used in discovery of biomarkers for early diagnosis of diseases such as cancer, neurological (Finehout *et al.*, 2007) and heart (Dubois *et al.*, 2011) and kidney (Wu *et al.*, 2010) just to mention a few. The technology has also been successfully applied in host-pathogen interaction studies in various crop plants including wheat (Wang *et al.*, 2005; Yang *et al.*, 2010), canola (Sharma *et al.*, 2007; Liang *et al.*, 2008; Marra *et al.*, 2010), potato (Liu and Halterman, 2009), rice (Konishi *et al.*, 2001; Mahmood *et al.*, 2006; Lin *et al.*, 2008; Yu *et al.*, 2008; Kim and Lee, 2009; Bhadauria *et al.*, 2010; Chi *et al.*, 2010), bean (Lee *et al.*, 2009) and apple (Gau *et al.*, 2004).

Thus the application of proteomics would greatly contribute in elucidating the effect of *Venturia inaequalis* have on the host plant (*Malus* x *domestica* Borkh.) at a very early stage of infection.

1.7.2 Transcriptomics Approach

The transcriptome of a cell or organism can be described as the set of all types of RNA molecules, whether coding messenger RNA (mRNA) or noncoding (ncRNA) including ribosomal RNA (rRNA), transfer RNA (tRNA), short interfering RNA (siRNA), micro RNA (miRNA), small nucleolar (snoRNA), or Piwi-interacting RNA (piRNA) expressed in a cell or whole organism (Martínez-Gómez *et al.*, 2011). Furthermore, it can also refer to the total set of genes that are actively expressed at any given time, responding to external and internal stimuli.

From the very first attempts to understand or study transcriptomes, Northern blot analysis, a candidate gene-based method (Alwine *et al*., 1977), was utilized. However this method was limited by the requirement of large amounts of RNA and the use of radioactivity. This issue was addressed a decade later with the development of the reverse transcription quantitative PCR (RT-PCR; Becker-Andre and Hahlbrock, 1989), which required less input RNA. Microarrays followed soon with the first publication to appear in 1995 (Schena *et al*., 1995). This method allowed the characterisation of expression levels of thousands of known or putative transcripts. Due to its versatility, it led to a multitude of expression-profiling initiatives, which aimed to comprehensively characterise expression signatures of different cell types and disease states in all types of organisms including bacteria, yeast, mammals, plants and fungi. This method gained such momentum and popularity that it became the method of choice for more than a decade (Mcgettigan, 2013) and its still in use today. This method is however not without drawbacks; it lacks the ability to detect novel transcripts and to study the coding

sequences of detected transcripts. Furthermore, the hybridisation-based nature of the method results in data received to be noisy, which interferes with reproducibility and cross-sample comparisons. As an alternative to microarrays, sequencing-based approaches to study transcriptomes were introduced with the development of sequencing technologies. This technology had an advantage over microarrays in that it has the ability to directly determine the identity and the abundance of transcripts such as in RNA-Seq. Transcriptome studies have evolved over the years, from sequencing individual cDNA clones to more complete portions of the transcriptome such as the full-length complimentary DNA (FLcDNA) method (Seki *et al*., 2002). A limiting factor for these sequencing-based methods was the high cost of the Sanger method (Sanger *et al*., 1977), as a result sequencing data suffered low coverage and it was insufficient to comprehensively characterise whole transcriptomes of multicellular species.

The development of expressed sequence tags (ESTs) in 1991 (Adams *et al*., 1991) addressed the cost limitation of Sanger sequencing to some extent. The first transcriptome profiling experiment using ESTs was on mammalian tissue. It was a less complete, less accurate but a cheaper alternative to the detection of expressed transcripts. It was still hampered by the expensive sequencing cost of the Sanger method and due to its labour intensiveness; it could not be routinely used on a transcriptome-wide scale. Furthermore, the low redundancy of sequencing reads, EST data were not suitable for estimating transcript abundance.

Serial analysis of gene expression (SAGE) was one of the first methods to be used to quantify gene expression on a global basis (Velculescu *et al*., 1995). It facilitated the use of Sanger sequencing for gene expression profiling. This method could address the shortcomings experienced by microarrays *viz.* the ability to detect novel transcripts, the ability to obtain direct measures of transcripts abundance, allow easier comparison between multiple samples, discovery of novel alternative spice isoforms. This method also suffered limitations such as a labour intensive cloning procedure, costly, and it produced very short sequences tags (14-21 bp), which were difficult to resolve for transcripts with similar coding sequence.

With the advancement of technology, sequencing has undergone a revolution with the sequencing of short-reads known as next-generation technology. The next generation sequencing platforms to enter the market were Roche/454, Illumina, Applied Biosystems SOLiD and Helicos HeliScope, had a much higher throughput than with Sanger sequencing and much cheaper. The first RNA-Seq paper was published in 2006 utilising 454/Roche technology (Bainbridge *et al*., 2006). This was already a major improvement on microarrays since it comprised 200 000 reads of 110 bp length, a total of 20 Megabases (Mb) of data. From 2008 onwards more papers were published using next generation sequencing such as Illumina. The very first Illumina machines generated 1 Gigabase (Gb) per run of data in 2006 rising to 1 Terabases (Tb) of data per run on the HiSeq2500 as of 2014 (http://www.illumina.com/systems/hiseq_2500_1500.ilmn). The emergence of next-generation sequencing has greatly benefitted the science community by reducing the cost of sequencing immensely, reduce experimental complexity, offering

improved transcript coverage and furthermore, made sequencing-based transcriptome analysis more readily available for individual laboratories. This technology has opened the door to explore new avenues and enable many new applications to be introduced for the study of transcriptomes.

1.7.2.1 RNA-Sequencing (RNA-Seq)

With the advancement of technology and more specifically the development of ultra-high-throughput sequencing technology, the measurement of transcriptome composition, discovery of new genes and exons are made possible. RNA-Seq is an alternative to microarrays and it even supercedes SAGE (Velculescu *et al*., 1995) and massively parallel signature sequencing (MPSS; Baginsky *et al*., 2010). What makes this technique even more attractive is the fact that it does not need prior genome annotation and it does not lean towards any biases introduced such as hybridisation and cross-hybridisation as is the case of microarrays (Baginsky *et al.*, 2010). The technique called RNA-Seq is a simple technique that avoids the use of bacterial cloning of cDNA input (Mortazavi *et al.*, 2008). This technique involves the direct sequencing of complementary DNAs (cDNAs) using high-throughput DNA sequencing technologies (Nagalakshmi *et al.*, 2010). Sequence reads are individually mapped to the organism's genome, counted to obtain the number and density of reads corresponding to RNA from each known exon, splice event or new candidate gene. The presence and amount of each RNA, is calculated and subsequently compared with the amount in any other sequenced sampled. If more than 40 million reads are collected for a sample, it should than be possible to detect and quantify RNAs from all biologically relevant abundance classes. The RNA-Seq technology

provides a more comprehensive understanding of complex eukaryotic transcriptomes, allowing for the identification of exons and introns. It allows for the mapping of boundaries of exons and introns, the identification of the 5´ and 3´ ends of genes. Furthermore, it allows for the identification of transcription sites, the identification of new splicing variants, and the monitoring of allele expression. The precise quantification of exon expression and splicing variants has been published in a number of papers to proof the uniqueness of this technology. However, RNA-Seq is also posed with challenges such as (i) how uniformly sequences from the entire transcript will be presented, (ii) quantification of sequence i.e. how much is needed to reliably detect and measure the concentration of RNAs of lower abundance classes and (iv) how the data will be quantified and how relative quantification will be converted to absolute RNA concentrations (Mortazavi *et al.,* 2008). Furthermore, transcriptomes of larger organisms containing genes with complicated structure also present special challenges. However, this technique becomes the method of choice in projects studying non-model organisms and the discovery of transcripts and genome annotation. The past years have seen a steady incline in researchers adopting the RNA-Seq technology in particular host-pathogen interaction studies such as apple (Gusberti *et al.,* 2013), rice (Kawahara *et al.*, 2012), black pepper (Gordo *et al.,* 2012), peach (Socquet-Juglard *et al.*, 2013), lettuce (De Cremer *et al.*, 2013), cucumber (Savory *et al.*, 2012), necrotrophic plant pathogen *Pythium ultimum* (Lévesque *et al.*, 2010) and apple scab transcriptome (Thakur *et al.*, 2013).

1.7.2.2 Illumina (Sequencing by synthesis)

Work done by Turcatti *et al*. (2008) on benzene-1,3,5-triacetic acid (BTA) and reversible deoxynucleotide terminators led to the development of the Illumina platform. Adaptors are attached to the many fragmented DNA strands and the molecules subsequently bend over and "bridge" hybridise to complementary adapters, thereby forming the template for the synthesis of their complementary strands. During this bridge PCR amplification stage, forward and reverse primers are attached to a glass surface, which allow the amplified constructs from a single template to be grouped into a cluster. *Bst* polymerase extends the template sequence during each step in the reaction and formaldehyde is used to denature the double stranded sequence (Turcatti *et al*., 2008). Following the amplification step, a flow cell with more than 40 million clusters is produced, wherein each cluster is composed of approximately 1000 clonal copies of a single template molecule. The amplification process results in several million clusters amplified within each lane of the flow cell (Turcatti *et al*., 2008). Following cluster construction, the double stranded amplified constructs are denatured into single strands. A sequencing primer is hybridised to the adaptor. The template is extended with a modified deoxynucleotide base, which is modified in two ways as (i) a reversible terminator and (ii) it is fluorescently labelled to correspond to the four nucleotide bases. After incorporation of the modified deoxynucleotide base on the sequencing strand, chemical cleavage is needed to remove the 3´-hydroxyl position, and the attached fluorescent molecule again starts a chain of reactions ending in the emission of a light signal. A charge coupled device (CCD) camera captures the signal and the incorporated base is then computationally determined in downstream analysis of the images. The proprietary Illumina analysis tools *viz.* Firecrest

and Bustard are used to analyse the raw data. The array is then prepared for the next cycle of base incorporation by enzymatically removing the blocking position of the incorporated base, and the next round of bases are flushed over the array. At every cycle of the sequencing process, only one base can be incorporated on the sequencing strand resulting in synchronous probe sequencing. At present, read lengths of up to 300 bp are possible, but there is a drop in quality of the reads as the read reaches the maximum read length. Quality of bases varies from greater than 75% to greater than 90% of bases above Q30 for single and paired-end reads (http://www.illumina.com/systems/hiseq_2500_1500.ilmn).

The development of the paired-end protocol, where both ends of the amplicons are sequenced, together with the extremely high-throughput (500 Gbp on a single flow cell with up to 1 Terabyte on dual flow cell) on the HiSeq 2500 platform, has made this technology ideal for genome re-sequencing and transcriptome studies where the digital expression on a specific transcript can be measured (http://www.illumina.com/systems/hiseq_2500_1500.ilmn). Typically, the 1G genome analyser from Illumina, Inc., is capable of generating 35-bp reads and producing at least 1 Gb of sequence per run in 2–3 days. With the HiScan SQ machine different read lengths were produced i.e. of 1 x 35 bp with output of 23-26 Gb would take approximately 1.5 days, whereas a 2 x 50 bp read length would give 67-75 Gb in approximately 4.5 days. For a longer read of 2 x 100 bp it would take approximately 8.5 days and output of 135-150 Gb (http://www.illumina.com/systems/hiscansq.ilmn). The HiScan SQ has been discontinued since.

1.7.3 RNA-Seq data analysis workflow

Transcriptomics is the study of the transcriptome i.e. identifying the full transcripts including large and small RNAs, novel transcripts from un-annotated genes, spicing isoforms and gene-fusion transcripts (Martin and Wang, 2011). A few years ago, a method developed by Wang *et al.* (2009), called RNA-Seq has revolutionised the way the transcriptomics are performed. Commonly, the term RNA-Seq refers to the dramatic increases in throughput of cDNA sequences obtained by the application of second generation sequencing platforms as discussed in section 1.7.2.1 (Wang *et al.*, 2009). A typical transcriptome assembly experiment involves the following steps *viz.* a library construction, data processing and assembling step. The library construction involves the purification of total RNA or mRNAs, fragmentation and conversion into a library of cDNAs containing sequencing adaptors. The cDNA is sequenced using next-generation sequencers such as Illumina HiSeq, SOLiD or Roche/454 to produce millions to billions of short reads from one end or both ends (paired-end) of the cDNA fragments. In the data processing phase the short reads are filtered and trimmed to remove sequencing errors and adaptors. These are now subsequently assembled to reconstruct the original RNAs and to assess their abundance (Martin and Wang, 2011).

1.7.3.1 Library construction

Preparation of cDNA libraries involves enriching for mRNA and during this step rRNA and abundant transcripts are removed. Poly(A) tail selection is done to extract mRNAs however, this approach will miss non-coding RNAs and mRNAs without poly(A) tails. cDNA is synthesised from the poly(A) tail mRNA and these fragments are fragmented

using restriction enzymes. Adapters are ligated and the adaptor-ligated cDNA is sequenced using high-throughput sequencing instruments.

1.7.3.2 Pre-processing of Data

In order to improve read quality and subsequent accuracy and computational efficiency of the assembly, artefacts need to be removed from the RNA-Seq data sets. There are three types of artefacts i.e. sequencing adaptors, low-complexity reads and duplicates. Several tools exist which can be employed such as FASTX Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/commandline.html), cutadapt (http://code.google.com/p/cutadapt/) and Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic) for removal of adaptor sequences. The latter can be used for an adaptor, low-complexity reads and sequencing error removal from single and paired-end format reads. Removal of these artefacts improves assembly quality and an added advantage is that it decreases the amount of computational power needed.

1.7.4 From RNA-Seq data to differential expression (DE) analysis

A typical RNA-Seq pipeline for DE analysis involved the following: (i) read mapping to the reference genome, (ii) mapped reads for each sample are assembled into gene-level, exon-level or transcript level expression summaries, (iii) normalisation of the summarised data in concert with the statistical testing of DE resulting in a ranked list of genes with associated *P*-values and fold changes and (iv) gaining biological insight by performing systems biology approaches.

1.7.4.1 Mapping

The filtered RNA reads discussed in section 1.7.3.2 can now be used to compare expression between samples for example infected versus uninfected tissue. The millions of short reads obtained need to be turned into a quantification of expression. To achieve this, read mapping or alignment to a reference genome, called reference based or *ab initio* strategy (Martin and Wang, 2011) needs to be done.

Mapping involves finding the unique location where a short read is identical to the reference genome. For the mapping step, splice-aware aligner software such as BLAT (Kent, 2002), TopHat (Trapnell *et al.*, 2009), SpiceMap (Au *et al.*, 2010), MapSpice (Wang *et al*., 2010) or GSNAP (Wu and Nacu, 2010) are used. Most of the short-read aligners make use of a first pass 'heuristic' match strategy, which quickly finds a reduced list of possible locations, followed by a thorough evaluation of all candidate alignments by a complex 'local alignment' algorithm (Oshlack *et al*., 2010). Two types of aligners exist *viz.* hash tables and Burrows-Wheeler Transform (BTW) (Li and Durbin, 2009) such as Bowtie (Langmead *et al*., 2009), Burrows-Wheeler alignment BWA (Li and Durbin, 2009) and SOAP2 (Li *et al*., 2009). Hash-table aligners can be extended but at an ever-increasing computational requirement, which make them less attractive to use. BTW aligners such as Bowtie (Langmead *et al*., 2009) on the other hand, map reads that closely match the reference very efficiently. Mapping of single reads often result in multimaps, which are discarded by most of aligners or allocated randomly or allocated based on an estimate of local coverage. Mapping of paired-end reads seams to alleviate this problem since both ends of the cDNA fragment map on the transcriptome. The advantage of

mapping short reads to a reference genome is that annotation of these reads is therefore not biased towards any known annotation.

1.7.4.2 Summarisation

Following alignment of reads, the genomic locations for the reads have to be summarised and a biologically meaningful unit has to be ascribed such as exons, transcripts or genes (Oshlack *et al*., 2010). The process of exon and gene specification is called annotation. Different strategies were developed to summarise the genomic locations. Mortazavi *et al.* (2008) employed a summarisation strategy, which counted the number of reads overlapping the exons in a gene. This was obtained by using a program called Enhanced Read Analysis of Gene Expresson (ERANGE) (Mortazavi *et al*., 2008). Other summarisation strategies include (i) the inclusion of reads along the whole length of the gene and thereby incorporating reads from introns (Oshlack *et al*., 2010), (ii) inclusion of reads that only map to coding sequences or (iii) *de novo* summarisation from predicted exons using Cufflinks (Trapnell *et al*., 2009). The choice of summarisation method has the potential to change the count for each gene dramatically.

1.7.4.3 Normalisation

Normalisation refers to the accurate comparison of expression levels between and within samples (Sultan *et al*., 2008). This is a crucial step in the analysis of differential expression from RNA-seq data. Once the reads have been summarised at the exon or gene level, the expression levels of each gene relative to other genes have to be

normalised. Transcript length vary from one to another, as a result, longer transcripts will have higher read counts as shorter transcripts. In such a case, normalisation is performed with the division of the summarised counts by the length of the gene. Mortazavi and his group (2008) came up with reads per kilobase of exon model per million (RPKM) as a means to normalise within-sample comparisons. This method accounts for both library size i.e. number of reads mapped and the effect of gene lengths (Mortazavi *et al*., 2008). A normalisation method developed by Marioni and his group (2008), adjusts by the total number of reads in the library. It accounts for the fact that more reads will be assigned to each gene if a sample is sequenced to a greater depth (Marioni *et al*., 2008). R statistical software packages such as edgeR (Robinson *et al*., 2013) and limma/voom (Smyth, 2004) allow normalisation of count table data.

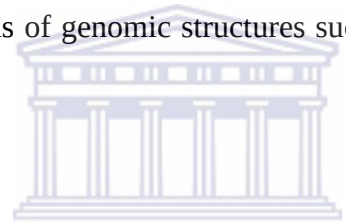
1.7.4.4 Differential expression (DE)

Differential expression analysis aim to highlight genes, which were changed significantly in abundance due to internal or external stimuli for example a pathogen infection in plants. This is usually done by taking the table of summarised counts for each library as discussed in 1.7.4.2 and performing statistical testing between the samples of interest. Since microarrays have dominated the field of gene expression studies for over a decade before the arrival of RNA-seq technology, many statistical methods were developed to analyse microarray data. Microarray studies were practically straightforward compared to RNA-se analysis, since microarray intensities are log-transformed and fixed in advance. Only the abundances of the specific RNA sequences that correspond to probes on the microarray are measured. As a result, quantification of differential expression is relatively

straightforward; measurements from the same probe were compared across samples. The data were analysed as normally distributed random variables, whereas RNA-seq data are summarised into count tables. For that matter, statistical models had to be developed to extract the most information form the RNA-seq data. Poisson distribution (Wang, *et al*., 2009) was initially proposed since it formed the basis for modelling RNA-seq count data. However, this method was inappropriate since it could not account for biological variability. In cases of biological replicates, the method was prone to high false positive rates. Despite the ideal results obtained from the RNA-seq platform such as low background and high sensitivity, biological replicates are still critical for identifying changes in RNA abundance in the samples and/or population being sampled.

Statistical methods such as edgeR (Robinson *et al*., 2010) were developed to analyse SAGE data. This method dealt with biological variability in samples/populations and has been applied to RNA-seq data recently. The issue of biological variability was addressed by using negative binomial distribution. A number of statistical models have been proposed including weighted likelihood (Robinson and Smyth, 2007), empirical estimation of the mean-variance relationship (Anders and Huber, 2010), an empirical Bayesian implementation (Hardcastle and Kelly, 2010) to address the issue of biological variability. Srivastava and Chen proposed an extension of Poisson distribution model (Srivastava and Chen, 2010) to improve the analysis of RNA-seq data and even a two-stage Poisson model, which tested for differential expression in two modes.

In RNA expression studies where biological replicates are limited, methods such as edgeR (Robinson *et al*., 2010), differential expression analysis of count data (DESeq) (Simon Anders and Huber, 2010), limma (Smyth, 2004) and later versions of Cuffdiff (Trapnell *et al*., 2010), are able to model the count variance across replicates as a non-linear function of the mean counts using parametric approaches such as normal and negative binomial distributions. However, the importance of biological replicates cannot be overemphasized since it provide the only measure of intrinsic, non-technical transcript expression variability and is critical for DE analysis. Other methods exist such as Alexsa (Griffith *et al*., 2010), DEXSeq (Anders *et al*., 2012) and DSGseq (Wang *et al*., 2013), which focus on the DE analysis of genomic structures such as splicing or transcriptional differences.



However these methods were not short of limitations. A recent method developed by Frazee *et al.* (2014) seems to address these shortcomings. They developed a method in which RNA-Seq data is analysed at single-base level (Frazee *et al*., 2014). It works by identifying differentially expressed regions (DERs) of interest by assessing the differential expression at each base of the genome. The genome is then segmented into regions comprised of bases, which show a similar DE signal. A measure of statistical significance is then assigned to this region. The DERs are then annotated using a reference database of genomic features. In a study where DER finder is compared to edgeR, DESeq and Cuffdiff, the authors found DER finder to outperform Cuffdiff and performed comparably with edgeR and DESeq. This type of method addresses the need to thoroughly extract all the information from RNA-seq data.

1.7.4.5 Systems Biology Approaches

Differential expression analysis often results in a long list of DE genes, which needs to be functionally annotated to give biological insight into the experiment. The expression changes of sets of genes can reveal a lot of biological information. These include biological processes, metabolic pathways, transcriptional programs and stress responses. Since microarrays have been mastered over the years, many tools were developed which focussed on gene set testing, network inference and knowledge databases in order to analyse DE lists obtained from microarray datasets. These tools include Gene Set Enrichment Analysis (GSEA) (Subramanian *et al*., 2005), Database for Annotation, Visualisation, and Integrated Discovery (DAVID; Dennis *et al*., 2003) and Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisa & Goto, 2000). With RNA-seq data analysis, the researcher is faced with biases, which are not present in microarray data. Transcript length biases as such will affect the analysis if the sets of genes have predominantly short of long genes. For instance, longer or highly expressed transcripts will have higher read counts and are more likely to be detected as differentially expressed compared with the short and/or lowly expressed counterparts. This ultimately results in greater statistical power to detect DE for long and highly expressed genes. This phenomenon called selection bias cannot be removed with normalisation or re-scaling. Failure to account for this will lead to biased result, affecting downstream analysis such as testing Gene Ontology terms for enrichment among DE genes (Ashburner *et al*., 2000; Oshlack and Wakefield, 2009). Young and his group developed an application specifically to address the biasness experienced with other methods (Young *et al*., 2010). This method performs Gene Ontology analysis on RNA-seq data, it reduces the

complexity and highlight the biological processes in genome-wide expression studies. They also showed that the most significant categories identified in their study matched the known biology. In addition to these methods, commercial packages were developed such as QIAGEN's Ingenuity Pathway Analysis (IPA) platform (http://www.ingenuity.com/products/rna-seq-expression-analysis), to aid in RNA-seq expression analysis and to give biological meaning such as linking DE genes to biological processes, stresses, diseases and molecular interactions.

1.8 Aims and Objectives of Project

The aim of this research project was to identify the response proteins in the interaction between susceptible *Malus* x *domestic* cv Golden delicious and *Venturia inaequalis*. This requires transcriptome and proteome analysis be performed on infected and uninfected apple leaves and the findings obtained could lead to the discovery of novel apple resistance genes and proteins.

The first step was to genetically characterise the *Venturia inaequalis* strain. To achieve this, an *in vitro* fungal culture was established by isolating a single isolate from the infected leaf material and (ii) genetically identify *Venturia inaequalis* by using SSR's as a means of confirming the identity.

For the transcriptomics section, the main objective was to inoculate apple seedling leaves, extract total RNA, construct cDNA libraries for RNA-seq, align RNA-seq data to the apple genome, extract and identify differentially expressed genes. In order to achieve the

goals set for this, bioinformatics tools were utilised to analyse the RNA-seq data, annotate and identify the differentially expressed genes and to biologically interpret the data.

For the proteomics approach, the main objective was to extract, separate and compare stress response proteins in the leaf proteome of infected and uninfected leaves. In order to achieve this objective, software performing comparative spot analysis and bioinformatics tools was used to identify the proteins of interest.

In the integrative study, the objective was to identify overlapping proteins in the Day 4 proteome and transcriptome. This was achieved by performing a comparative study of the up-regulated proteins and the up-regulated genes for Day 4.

Chapter Two - Materials and Methods

2 Materials and Methods

2.1 General chemicals and enzymes

All chemicals used in the study and their suppliers are listed alphabetically in Table 2.1.

Table 2.1 List of chemicals

Chemical	*Supplier*
Acetone	Merck
Acetonitrile	Merck
40% Acrylamide-Bis Solution, 37.5:1 (2.6%)	BIO-RAD
Agar bacteriological	Merck
Agarose D1 LE	Whitehead Scientific
Ammonium hydrogen bicarbonate	Merck
Ammonium persulphate (APS)	Sigma
100x Bio-Lyte 3/10 Ampholyte	BIO-RAD
Boric acid	Merck
BIO-RAD Protein Assay dye reagent concentrate	BIO-RAD
Bovine Serum Albumin Fraction V (BSA)	Roche
Bromophenol Blue sodium salt	Sigma

Buffer saturated phenol	Sigma
3-[(3Cholamidopropyl)dimethylammonio]-1propanesulfonate (CHAPS)	Sigma
Chloroform	Merck
Coomassie Brilliant Blue R-250 (CBB)	BIO-RAD
Diethylene pyrocarbonate (DEPC)	Sigma
Disodium hydrogen phosphate 2-hydrate	Merck
Dithiothreitol (DTT)	Fermentas
dNTPs	Sigma
Ethylene diamine tetra-acetic acid (EDTA)	Merck
Ethanol	Kimix
Ethidium bromide	Merck
Formaldehyde	Merck
GeneScan-500 LIZ™ size standard	AB Sciex
Glacial acetic	Merck
Glycerol	Merck
Glycine	BIO-RAD
Hi-Di Formamide	AB Sciex
Hydrochloric acid (HCl)	Merck
Iodoacetamide	BIO-RAD
Isoamylalcohol	Merck
Isopropanol (propan-2-ol)	Merck
Kapa PCR Kit	KapaBiosystems

Lithium Chloride	Merck
ß-mercaptoethanol	Sigma
Methanol	Merck
Mineral oil	BIO-RAD
3-[N-Morpholino]propanesulfonic acid (MOPS)	SIGMA
Oxy-tetracycline	Sigma
PageRuler™ unstained protein ladder	Fermentas
Potato dextrose broth	Saarchem
Potato dextrose agar (PDA)	(BD) Becton and Dickinson Company
Protease inhibitor cocktail for fungi and yeast	Sigma
Polyvinylpyrolidone (PVP 40)	Sigma
Polyvinylpolypyrolidone (PVPP)	Sigma
RNAse A	Roche
RNase AWAY®	Molecular BioProducts
Sequazyme™ Peptide Mass Standards Kit	AB Sciex
Sodium acetate	Saarchem
Sodium chloride	Merck
Sodium dodecyl sulphate (SDS)	BIO-RAD
Sodium hydroxide	Merck
Sodium sulphite	NT Laboratory Supplies

Sodium hypochloride solution (12.5%)	Kimix
Sucrose	Merck
N,N,N^1,N^1-tetra methylethylene-diamine (TEMED)	BIO-RAD
Thiourea	Sigma
Trichloroacetic acid (TCA)	Merck
Tris(hydromethyl)aminomethane (Tris)	BIO-RAD
Triton-X 100 (iso-octylphenoxypoly-ethoxyethanol)	BDH
Trypsin	Promega
Tween® 20 (polyoxyethylene 20 sorbitan)	Merck
Urea	Sigma

All the chemicals used were of AnalaR grade. All buffers were prepared with dH_2O purified using the Maxima water purification system (Elga, Wycombe, England).

2.2 Stock Solutions and Buffers

80% Acetone: 80% (v/v) Acetone in distilled water

1% Agarose gel: 1% (w/v) agarose in 1x TBE buffer

0.5% Agarose sealing solution: 0.5% (w/v) agarose prepared in 1X SDS-PAGE running buffer with a tint of bromophenol blue

1% Agarose: 1% (w/v) agarose in 1× TBE

10% APS: 10% (w/v) APS in distilled water. The solution was freshly prepared before use.

Bradford reagent: 1 part BIO-RAD Protein assay dye reagent concentrate diluted with 4 parts distilled water

Bufffer I: 10 mM potassium-phosphate buffer (pH 7.4) containing 0.07% (w/v) DTT, 1 complete™, EDTA-free (protease inhibitor cocktail tablet) (Roche) and 40 µl/ml protease inhibition cocktail for fungi and yeast

Buffer II: 10% (w/v) TCA in 100% (vv) acetone containing 0.07% (w/v) DTT

Buffer III: 100% (v/v) acetone containing 0.07% (w/v) DTT

Buffer IV: 8 M urea, 2% (w/v) CHAPS, 20 mM DTT and 0.5% (v/v) ampholyte in distilled water

5mg/ml BSA stock solution: 5 mg/ml BSA in urea buffer

CBB staining solution I: 50 ml of 1.25% (w/v) CBB stock solution and 10% (v/v) glacial acetic acid and 25% (v/v) propan-2-ol in distilled water

CBB staining solution II: 6.25 ml of 1.25% (w/v) CBB stock solution and 10% (v/v) glacial acetic acid and 10% (v/v) propan-2-ol in distilled water

CBB staining solution III: 6.25 ml of 1.25% (w/v) CBB stock solution and 10% (v/v) glacial acetic acid in distilled water

0.1% DEPC water: 1 ml of DEPC in 1000 milliliter distilled water

Displacing solution: 0.375 M Tris-HCl, pH 8.8, 50% (v/v) glycerol with a tint of bromophenol blue

6x DNA Loading buffer: 0.25% (w/v) bromophenol blue, 0.25% (w/v) xylene cyanol FF, 5 mM EDTA and 50% (v/v) glycerol in distilled water

Equillibration buffer I: 8 M urea, 2% (w/v) SDS, 0.05 M Tris-HCl, pH 8.8
and 20% (v/v) glycerol and 1% (w/v) DTT

Equilibration buffer II: 8 M urea, 2% (w/v) SDS, 0.05 M Tris-HCl, pH 8.8 and 20% (v/v) glycerol and 2.5% iodoacetamide

Ethanol: 99.99% **(**v/v) Absolute ethanol

70% Ethanol: 70% (v/v) absolute ethanol and 30% dH_2O (v/v)

1x FA gel buffer: 20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA (pH 7) in distilled water

0.1 M HCl: 0.1 M HCl in distilled water

IEF Rehydration buffer: Urea buffer (7M urea, 2 M thiourea, 4% CHAPS) and 0.2% DTT and a tiny pinch of bromophenol blue

8 M LiCl: 8 M LiCl in distilled water

Lysis buffer (i) for genomic DNA extraction from fungus: 40 mM Tris-acetate, 20 mM sodium acetate, 1 mM EDTA, 1% (w/v) SDS pH 7.8 and 0.6% (w/v) sodium sulphite in distilled water

Lysis buffer (ii) for total RNA extraction from apple leaves: 100 mM Tris-HCl pH 8.0, 600 mM NaCl, 20 mM EDTA, 4% (w/v) SDS and 0.06% sodium sulphite in distilled water

Lysis buffer (iii) for total RNA extraction from apple scab: 100 mM Tris-HCl pH 8.0, 600 mM NaCl, 20 mM EDTA, 4% (w/v) SDS in distilled water

Na-acetate (pH5.2): 3 M Na-acetate in distilled water

5 M NaCl: 5 M NaCl in distilled water

80% (v/v) Methanol containing 0.1 M ammonium acetate: 80% (v/v) Methanol in distilled water with 0.1 M ammonium acetate

PCI (Phenol:Chloroform:Iso-amylalcohol): Buffer saturated Phenol 50% (v/v) : Chloroform 48% (v/v) : Iso-amylalcohol 2% (v/v)

RNAse A: 0.0625 mg.ml^{-1} RNAse A in 2x RNAse buffer

5x RNA loading buffer: 0.16% (v/v) saturated aqueous bromophenol blue, 4 mM EDTA, pH8.0, 089 M Formaldehyde 37% (v/v), 20% (v/v) Glycerol, 4x FA gel Buffer, 7.71 M Formaldehyde

10% SDS: 10% (w/v) SDS in distilled water

SDS Buffer: 30% (w/v) Sucrose, 2% (w/v) SDS, 0.1 M Tris-HCl pH 8.0, 5% ß-mercaptoetanol

5x SDS electrophoresis buffer: 0.1% (w/v) SDS, 250 mM glycine, 25 mM Tris, pH 8.3

2x SDS sample loading buffer: 60 mM Tris-HCl pH 6.8, 2% (w/v) SDS, 10% (v/v) glycerol, 200 mM DTT, 0.025% (w/v) bromophenol blue

1x SDS-PAGE running buffer: 25 mM Tris, 192 mM glycine containing 0.1% (w/v) SDS

1 x TBE: 10% (v/v) of 10× TBE (0.9M Tris, 0.89 M boric acid, 0.032 M EDTA pH 8.3) in 1 liter of distilled water

1X TE buffer: 1 mM EDTA, 10 mM Tris-HCl, pH 7.8 in distilled water

0.5 M Tris-HCl (stacking gel buffer), pH 6.8: 0.5 M Tris in distilled water adjusted to pH 6.8 with concentrated HCl

1.5 M Tris-HCl (separating gel buffer), pH 8.8: 1.5 M Tris in distilled water adjusted to pH 8.8 with concentrated HCl

Urea buffer: 7 M urea, 2 M thiourea and 4% CHAPS. The solution was kept at -20ºC

2.3 Plant material

Seeds obtained from Golden Delicious open pollinated population apples were stratified for 3 months at 4ºC and planted into seedling trays in a greenhouse. This was done at the ARC Experimental Farm, GPS coordinates (-33.843865, 18.978881) (http://www.onlinebrandambassadors.com/app/map/find-gps/), Bien Donné, Simondium, South Africa. The seedlings were planted January 2008 and were given high levels of nitrogen and balanced food growth for optimal growth. Three-week old seedlings were ready for inoculation. Leaf samples were collected and frozen in liquid nitrogen and stored at -80ºC. These were used for total RNA and protein extraction.

2.4 Characterisation of fungal strain

2.4.1 Microsatellite primers for fungal identification

Twenty-one microsatellite primer pairs published by Guerin *et al.* (2004) of *Venturia inaequalis* were synthesised at AB Sciex (Foster City CA, USA) and the primer closest to the repeat were labelled with one of the four fluorescent dye colours and fluorescently-labelled using the 6-FAM™ (blue), NED™ (yellow), PET™ (red) and VIC™ (Green) dyes.

2.4.2 Amplification of DNA using polymerase chain reaction (PCR)

A single locus screen of microsatellite markers were done in order to test the ability of specific primer pairs to amplify target DNA and generate amplification products or fragments.

2.4.3 PCR conditions for single locus screening

Polymerase chain reaction was performed with fluorescently labelled primers. Amplifications were done in volume of 25 µl containing 5 µl of 5x buffer from KapaBioscience, 0.75 µl of 50 mM magnesium chloride. 0.2 µl of dNTPs, 0.2 µl forward and 0.2 µl reverse primer, 0.125 µl of KapaTaq, 2 µl of 20 ng/µl genomic DNA and 16.525 µl of distilled water. The thermal cycling conditions were as follows: a denaturation step of 5 min at 95°C, 30 cycles consisting of 94°C for 30 sec, 58°C for 30 sec, 72°C for 40 sec, a final elongation step at 72°C for 10 min and 4°C hold. PCR amplifications for genotyping were performed with a 9700 Thermal Cycler AB Sciex (Foster City CA, USA). Alleles were scored against a fluorescently labelled size standard in an ABI 3700 automated sequencer (AB Sciex, Foster City, California).

2.4.4 Automated electrophoresis on ABI PRISM® 3130xl Genetic Analyzer

Electrophoresis was carried out according to manufacturer's instructions with 25 µl of formamide, 3 µl of PCR product and 0.1 µl of GeneScan 500 LIZ standard. Samples were loaded into the ABI PRISM® 3130xl (16-capillary array system) Genetic Analyzer AB Sciex (Foster City CA, USA). Size determination of 6-FAM™, NED™, PET™ and VIC™ labelled primes was done using an internal standard LIZ™ (GeneScan™ -500 LIZ™) fluorescent dyes, POP-7 sieving polymer matrix, 1x Genetic analyser buffer with EDTA and 36 cm x 50 µm uncoated capillaries. The mixture was denatured at 96°C for 5 minutes with snap cooling prior to loading them into the autosampler tray. Samples were injected for 15 s at 15 000 V and separated at 15 000 V for 24 min with a run temperature

of 60°C. The resulting data can be displayed as an electropherogram using GeneScan® software AB Sciex (Foster City CA, USA).

2.5 Infection and Inoculation of seedlings

This experiment was performed at the ARC Experimental Farm, Bien Donné, Simondium South Africa. In February 2008 the seedlings were inoculated with *Venturia inaequalis*. The inoculum, which is a mixture of different isolates, were collected by washing of the conidia spores from infected leaves obtained from the orchards at Bien Donné. The infected leaves were soaked for one hour in water and the inoculum concentration was optimised to 5 x 10^4 conidia/ml in water. Seedlings were then inoculated using a low-pressure spray gun in an infection chamber at 99% relative humidity (RH) and 16ºC. Seedlings for control were mock inoculated with distilled water being sprayed on. The seedlings were kept wet for 48 hours by applying a fine water spray for 20 seconds every 30 minutes. The seedlings were then transferred to a greenhouse at 20-25ºC and kept there for approximately 25 days until disease symptoms developed.

2.5.1 Total number of seedlings

Ten trays (7 x 14 wells) from a Golden Delicious open pollinated population of seedlings were used in this experiment. Three sets of trays (3 trays x 3 replicates) were used for the treatment and 1 tray was used for the control. In the control tray (2 x 14 wells) x 3 replicates were dedicated for seedlings. Sampling was done in triplicate for control and treatment.

2.5.2 Collection of leaf material

Collections of samples i.e. control and treatment, were taken every second day, starting at Day 0. This was done until Day 12. Two young, fully developed leaves per individual plant were collected from the population. Samples were collected from the first two rows (2 x 14) for Day 0 from all the 3 trays. Collection of samples was done in triplicates. For Day 2, collection was from the next two rows. This pattern of collection was followed until Day 12. These were bagged and put immediately in liquid nitrogen and transported to the laboratory where they were stored at -80ºC. Collected sample material was used for total RNA and total protein extraction.

2.6 Establishment of an *in vitro* grown *Venturia inaequalis* culture from an isolate and culture conditions

Spores from a *Venturia inaequalis* isolate were obtained from ARC Experimental Farm, Bien Donné, Simondium, South Africa and germinated on water agar. Germinated spores were transferred to potato dextrose agar plates and left to grow at room temperature for a month. Mycelium plugs of 5 mm^2 in size were cut from the outer edge of the mycelium growth and transferred to 500 ml of potato dextrose broth containing 25 mg oxy-tetracycline and grown at 21ºC with agitation at 100 rpm for 2-3 months. Mycelia were harvested by transferring the cultures to 50 ml tubes and centrifuged at 6 000 x *g* for 30 minutes. The supernatant was discarded and the mycelia were rinse with distilled water and centrifuged again for 30 minutes. The water was discarded and the mycelia were lyophilised and stored at room temperature and -80°C.

2.6.1 Genomic DNA extraction from apple scab

Genomic DNA was extracted from *Venturia inaequalis* using a modified protocol (Al-Samarrai and Schmid, 2000). Two hundred milligram of freeze-dried mycelium was ground to a fine powder with a mortar and pestle in liquid nitrogen. Four milliliter of lysis buffer (i) (section 2.2) was added into the frozen mycelia and allowed to thaw. Five hundred microliter of homogenate was transferred to 2 ml Eppendorf tubes and lysed by pipetting with a Gilson P 1000 pipetman until the viscosity of the suspension was significantly reduced and the formation of froth indicated the detachment of DNA from polysaccharides. At least 30 cycles of vigorous pipetting was performed. Five microliter of RNAse A (section 2.2) was added and the mixture was incubated for 20 min at 37°C. To facilitate the precipitation of most polysaccharides, protein and cell debris, 165 µl of 5 M NaCl solution (section 2.2) was added and the components mixed by slowly inverting the tubes several times. The suspension was centrifuged at 16 000 x *g* for 20 min at 4°C, the supernatant was immediately transferred to fresh tubes and 400 µl of chloroform and 400 µl of phenol were added. The solution was mixed by gently inverting the tubes until the solution became milky. After centrifugation at 16 000 x *g* for 20 min, the aqueous phase was removed and extracted with an equal volume of chloroform. The DNA in the aqueous supernatant was precipitated with two volumes of ice-cold absolute ethanol. To free the DNA from polysaccharide the precipitate was resuspended in 500 µl of lysis buffer (i) (section 2.2) and gently mixed. Then 165 µl of 5 M NaCl (section 2.2) was added and the tubes were gently inverted several times. The suspension was then chloroform-extracted as described above. Usually, after centrifugation for 10 min at 16 000 x *g*, the aqueous phase was clear and the DNA was precipitated again with ice-cold

absolute ethanol. The DNA was pellet by centrifuging for 10 min at 16000 x *g*. The precipitated DNA was washed three times with 70% ice-cold ethanol, air-dried and dissolved in 50 µl of 1x TE buffer (section 2.2) and stored at -20°C.

2.7 DNA Electrophoresis

2.7.1 Agarose gel electrophoresis

Fungal genomic DNA was size separated by horizontal gel electrophoresis in agarose. A 1% agarose gel (section 2.2) was prepared by completely suspending 1 g of agarose in 100 ml of 1× TBE and melted in a microwave oven. This was allowed to cool to about 50ºC and ethidium bromide (0.5 µg/ml) was added and poured into a gel tray with a 8-10 well comb in place and allowed to set. Prior to loading the sample it was mixed with 2 µl of 6x DNA loading buffer (section 2.2) and loaded into the wells of a 1% gel and electrophoresed in 1× TBE (section 2.2). The voltage and lengths of time depended on the size of the fragments being separated (smaller fragments at low voltage and shorter time). The conditions most frequently used were 10 V/cm for 1-2 hours. The gel was visualised with UV transillumination using the Gel-Doc UV Transilluminator (BIO-RAD, Hercules, California, USA).

2.8 Total RNA extraction

2.8.1 Pre-Treatment of glass and plastic-ware

All glass and plastic-ware used in the extraction and handling of RNA were soaked overnight in 0.1% DEPC (section 2.2). These were afterwards sterilised by autoclaving at 121°C for 20 minutes before drying in an oven at 100°C. Pestles, mortars, Eppendorf tubes and micropipette tips were pre-treated with 0.1% DEPC for removal of RNase. Micropipettes and electrophoresis apparatus were wiped with RNase AWAY®.

2.8.2 Total RNA isolation from apple (*Malus* x *domestica*) leaves

The youngest leaves from seedlings from the greenhouse were collected and immediately frozen in liquid nitrogen. Total RNA was isolated following a modified protocol (Menhaj *et al*., 1999). The leaves were ground in liquid nitrogen and 15 ml lysis buffer (ii) and 15 ml PIC mix (section 2.2) were added. Following shaking of samples for 20 min, samples were centrifuged at 12 000 x *g* for 20 min at 4°C. Following that, 15 ml PIC mix were added to the supernatant and centrifuged for 15 min in the same conditions. The supernatant was transferred to a fresh tube and the RNA was allowed to precipitate in the presence of 3.5 M LiCl overnight at 4°C. After precipitation, the pellet obtained was dissolved in 1 ml DEPC treated water and precipitated by incubation for 1 hour at -20ºC in the presence of 500 µl 3 M Na-acetate (pH 5.2) and 5 ml of ice cold absolute ethanol. After washing in 70% (v/v) ice cold ethanol, the pellet was air dried, dissolved in 100 µl DEPC treated water and stored at -80ºC. All solutions, except lysis buffer and PCI, were treated with DEPC water.

2.8.3 Denaturing agarose gel electrophoresis

Total RNA extracted from sample tissue was electrophoresed on denaturing agarose gel. A 1.2% denaturing agarose gel was prepared by melting 1.2 g of agarose in 100 ml of 1x FA gel buffer (section 2.2). The molten agarose was equilibrated to 65ºC in a water bath. Formaldehyde (37%) and ethidium bromide were added to a final concentration of 0.44 M and 0.2 mg/ml respectively and thoroughly mixed by swirling. The molten gel was cast and allowed to set at room temperature. The gel was pre-equilibrated in 1x FA gel buffer for 30 min prior to running the electrophoresis.

2.9 Total protein extraction from plant tissue

2.9.1 Total plant protein isolation from apple leaves (*Malus* x *domestica*)

A modified protocol was used for the extraction of total protein from young apple scab infected and uninfected leaves (Wang *et al.*, 2006). Approximately 2 g of leaf tissue was ground to a fine powder in a motar and pestle under liquid nitrogen. Half a gram of PVPP and 5 ml TCA in acetone was added to the ground leaf tissue and homogenised. The homogenate was transferred to 2 ml Eppendorf tubes and centrifuged at 16 000 x *g* at 4ºC for 3 min. The supernatant was discarded and 80% (v/v) methanol containing 0.1 M ammonium acetate was added. This was mixed properly by vortexing. Centrifugation was performed at 16 000 x *g* at 4ºC for 3 min. The supernatant was discarded. The tube was filled with 80% (v/v) acetone, and vortexed until pellet is fully dispersed. It was centrifuged at 16 000 x *g* for 3 min at 4ºC. The supernatant was discarded and the pellet

was air-dried for 10 min to remove residual acetone. Phenol and SDS buffer in a 1:1 ratio was added, i.e. 500 µl phenol and 500 µl SDS buffer. This was mixed thoroughly by vortexing and incubated for 5 min at room temperature. This was followed by centrifugation at 16 000 x *g* for 3 min at 4ºC. The supernatant was transferred to a fresh 2ml Eppendorf tube and filled with 80 % (v/v) methanol containing 0.1 M ammonium acetate. This mixture was allowed to precipitate by incubating at -20ºC overnight followed by centrifugation at 16 000 x *g* for 3 minutes at 4ºC. The supernatant was discarded and the pellet was washed once with 100% (v/v) methanol and once with 80% (v/v) acetone. During each wash step, mix thoroughly by vortexing the pellet and centrifuge as above. The pellet was air-dried and resuspended in urea buffer (section 2.2). The protein concentration was determined in the supernatant using the Bradford Assay (section 2.2) with bovine serum albumin (BSA) as standard. The sample was analysed on a 12.5% SDS-PAGE gel or stored at -80ºC for future use.

2.10 Protein Quantification

A modified method described by Bradford (1976) was used to determine total protein concentration for all samples (Bradford, 1976). In this method BSA standards were used in duplicate from a 5 mg/ml stock solution in 2 ml plastic cuvettes as indicated in Table 2.2. Extracted protein samples were also prepared in duplicate in 2 ml plastic cuvettes by mixing 5 µl of unknown protein sample with 5 µl of urea buffer, 10 µl of 0.1 M HCl and 80 µl of distilled water. The Bradford reagent was diluted five times with distilled water prior to use. 900 µl of the diluted Bradford reagent was added to all standards and protein extracts, mixed and incubated for 5 minutes at room temperature. Absorbance was

measured at 595 nm on Helios E spectrophotometer (Thermo Spectronic) using the 0 mg/ml BSA standard as a blank solution. The values obtained were used to plot a standard curve from which the concentrations of all the unknown extract samples were calculated.

Table 2.2 BSA standards preparation table for protein quantification

BSA Final Concentration (μg/μl)	*BSA 5 mg/ml stock solution (μl)*	*Urea Buffer (μl)*	*0.1 M HCl (μl)*	*Sterile dH_2O*	*Bradford Reagent (μl)*
0	0	10	10	80	900
5	1	9	10	80	900
10	2	8	10	80	900
20	4	6	10	80	900
40	8	2	10	80	900
50	10	0	10	80	900

2.11 One dimensional (1-DE) Sodium Dodecyl Sulphate-Polyacrylamide Gel electrophoresis (SDS-PAGE)

The proteins of apple leaves were separated respectively on 1-DE SDS-PAGE as described by Laemmli (1970). Two types of gels are used in this system *viz.* the stacking and the separating gel. The stacking gel has large pores that allow proteins to concentrate into thin zones before they enter the resolving gel. In this gel, protein mobility depends

on the net protein charge rather than the molecular weight. The separating gel is more compact than the stacking gel and proteins are separated on the basis of molecular weight (Westermeier, 2005).

For 1-DE gels, the Mini-PROTEAN® 3 Electrophoresis Cell gel casting system (BIO-RAD) was used. Separating gels of 12.5% (v/v) were prepared from a 40% Acrylamide/Bis stock solution 37.5: 1 (2.6% C) using 1.5 M Tris-HCl separating buffer, pH 8.8 as indicated in Table 2.3. The separating gel was poured first and overlaid with 1 ml of 100% isopropanol and allowed to polymerise. Then isopropanol was poured off and the separating gel was rinse thoroughly with distilled water and gel surfaces were dried off using filter paper. The stacking gels 5% (v/v) were prepared from a 40% Acrylamide/Bis stock solution 37.5: 1 (2.6% C) using 0.5 M Tris-HCl gel buffer, pH 6.8 as indicated in Table 2.3. The separating gel was overlaid with the stacking gel and the 1 mm (10-well comb) (BIO-RAD) was put in place.

Protein samples were mixed with an equal volume of 2× SDS sample loading buffer (section 2.2), boiled for 5 min at 95°C on a heating block (Dry Block Heater, FMH Instruments). Samples were centrifuged for a few seconds before loading on the gel. Total protein quantities of 20 µg per well were loaded onto the 10-welled gels and then electrophoresed in 1× SDS electrophoresis running buffer using a PowerPac™ Universal Power supply (BIO-RAD). The run was started with 100 V for the first 30 minutes and thereafter increased to 140 V. Electrophoresis was stopped when the dye front reached the bottom of the separating gel. The gels were removed from the plates, stained with Coomasie Brilliant blue and destained as described in section 2.12.7.

Table 2.3 Preparation of resolving and stacking gels for SDS-PAGE

	12.5% Separating Gel	*5% Stacking Gel*
Distilled Water	4.3 ml	1.48 ml
40% Acrylamide/Bis stock solution 37.5:1 (2.6% C)	3 ml	0.25 ml
1.5 M Tris-HCl, pH 8.8	2.5 ml	-
0.5 M Tris-HCl, pH6.8	-	0.25 ml
10% SDS	0.1 ml	0.02 ml
10% APS	0.1 ml	0.02 ml
TEMED	6 µl	2 µl
Total Volume	10 ml	2 ml

2.12 Second Dimension (2-DE) Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.12.1 Sample load

The optimal quantity of protein to load for electrophoresis varies widely depending on factors such as sample complexity, the length and pH range of the Immobiline DryStrip gel, and the method of visualizing the 2-D separation.

2.12.2 Sample preparation for 2-DE Gels

Total soluble proteins extracted from tissue *viz.* apple leaves (section 2.9) stored at -20°C were thawed on ice and protein concentrations were determined by Bradford assay section 2.10. The amount of protein loaded on a 2-DE gel is mixed with IEF rehydration buffer section 2.2 and 0.2% of ampholytes (pH 3-10) to a total volume of 125 µl (7 cm IPG strip). The sample was than vortexed and pulse centrifuged for a few seconds at 16000 x *g* at room temperature.

2.12.3 Rehydration of 7 cm IPG Strips

Total soluble protein extracts (300 µg) as prepared in section 2.12.2 was loaded onto the individual channels of an Immobiline™ Dry Strip Reswelling Tray (GE Healthcare, Amersham, UK). Linear, 7 cm ReadyStrip™ IPG strips of pH range 3-10 or pH range 4-7 (BIO-RAD) were carefully placed on top of the sample, gel side in contact with the sample, avoiding trapping of air bubbles. Each strip was covered with mineral oil to prevent dehydration or crystallisation of urea. This would allow passive rehydration of the gels to their original thickness of 0.5 mm for at least 15 hours.

2.12.4 First Dimension IEF of IPG Strips

After rehydration, IPG strips are removed from the gel tray and rinsed with distilled water to remove unabsorbed proteins and oil. Strips were carefully blotted with moist filter paper. Strips were placed gel side up on the focusing platform of the Ettan™ IPGphor II™ (GE Healthcare). Prior to mounting the electrode pads, wicks were moistened with

distilled water and were placed at the ends of both the anodic and cathodic ends of the IPG strips to collect excess salts and impurities from the sample during focusing. The IPG strips were covered with mineral oil to prevent sample evaporation and carbon dioxide absorption during focusing. Isoelectric focussing were executed in a three phase stepwise programme at 20°C as indicated in Table 2.4 for 7 cm IPG strips. Following isoelectric focussing, 7 cm IPG strips were equilibrated in preparation for the second dimension (section 2.12.5).

Table 2.4 Isoelectric focussing programme for 7 cm IPG strips

STEP	*Volts (V)*	*Duration (hrs)/Volt hours (Vhrs)*
1	200	0.10 hr
2	3 500	2800 Vhrs
3	3 500	3700 Vhrs

2.12.5 Equilibration of IPG Strips

Following IEF, the strips were equilibrated in SDS-containing buffers in order to solubilise focused proteins by coating them with SDS for separating on the basis of their mass. The equilibration process involves two-steps *viz.* the reduction of disulphide bonds and the alkylation of the resultant sulfhydryl groups of the cysteine residues. During the first equilibration step, strips were placed in an equilibration tank gel side up and incubated with 2 ml of equilibration buffer I (section 2.2) for 15 minutes with gentle shaking. The buffer was discarded prior to addition of 2 ml of equilibration buffer II (section 2.2) and incubation for 15 minutes with gentle shaking.

2.12.6 Second Dimension SDS-PAGE of Mini Format Gels

Mini format 2-DE SDS-PAGE gels were cast 10 cm (width) and 8 cm (height) spacer glass plates (BIO-RAD) mounted with 1 mm thick spacers using the Mini-PROTEAN® 3 Multi-Casting Chamber (BIO-RAD). Separating gel solution was prepared as described in Table 2.3. The gel was poured into the cast plates according to the instruction manual and overlaid with 100% isopropanol. The gels were allowed to polymerize followed by discarding of isopropanol and rinsing the gel surfaces with distilled water. Gel surfaces were blotted dry with filter paper to remove excess water.

The equilibrated 7 cm strips were rinsed 1x SDS-PAGE running buffer (section 2.2) and placed on top of the mini format 12.5% separating gel, overlaid with molten 0.5% agarose sealing solution (section 2.2). Prior to this 3 µl of PageRuler™ unstained protein ladder were spotted on small pieces of filter paper, air-dried and placed at the anodic side of each IPG strip. Electrophoresis was carried out using the Mini-PROTEAN® 3 Dodeca™ cell (BIO-RAD) at 100 V for the first 30 minutes and then at 140 V until the tracking dye reached the bottom of the glass plates. Gels were then stained with Coomassie Brilliant Blue (section 2.12.7).

2.12.7 Coomassie Brilliant Blue Staining of SDS-PAGE Gels

Total protein samples loaded on either 1-DE or 2-DE SDS-PAGE were detected by using a modified CBB R-250 staining protocol that uses a three sequential staining step. Following electrophoresis, the gels are placed in CBB staining solution I (section 2.2), heated for 1 min in a microwave oven and incubated for 30 minutes with shaking at room temperature. The staining solution was discarded and the process was repeated for CBB

solution II and III (section 2.2) as described above. Following the sequential staining process, the gels were destained by immersing in destaining solution (section 2.2) with shaking at room temperature until the protein bands or spots were clearly visible against a clear background. The gels were imaged using a Molecular Imager PharosFX Plus System (BIO-RAD).

2.13 Comparative analysis

2.13.1 Comparative analysis using Melanie software 7.0

Comparative analysis of 2-DE SDS-PAGE gels of Day 4 Control and Treatment was carried out using Melanie 7.0 (Genebio, Geneva, Switzerland). Two-dimensional gels were initially imaged using the Molecular Imager PharosFX Plus System (BIO-RAD) and then analysed according to Melanie 7.0 user manual. All the experiments analysed with this software had three biological and three technical repeats for each treatment. In order to find corresponding spots between different experiments, hierarchical match structures were created. Gels or matched sets used as reference are marked red. Only spots matched with a spot in another gel are included in the Gel and Class Analysis Tables. In order to continue with the analysis, match sets Control (C) and Treatment (T) are merged to create the root match set CT. This root match set contains the folders Match and Classes. Once the match set has been set up, spot detection can be performed by adjusting the parameters *viz.* Smoothness, Saliency and Minimum volume. A landmark, that is a clearly defined corresponding spot in each gel, can be added. Gels are matched and corresponding spots that are mismatched can also be edited. In order to find protein expression variations between the Control and the Treatment, classes are defined. This

would allow you to create tables, graphs and apply statistics to the spots and classes. For statistical analysis, the following were selected *viz.* Central tendency (Mean), Dispersion (Mean squared deviation) and Student's *t*-test. A Class Analysis Table with all the matching spots was created. By selecting a spot from the Class Analysis table, a histogram is obtained to give an indication of fold expression. The data obtained from the Class Analysis Table i.e. the Student's *t*-test values (95% significance level) are used to calculate the expression fold change between the Control and the Treatment for the matched spots. Protein spots of interest were picked for identification using mass spectrometry and database searching (section 2.13.2.2).

2.13.2 Mass spectrometry (MALDI TOF) analysis

The Central Analytical Facility (CAF), Proteomic laboratory, at the University of Stellenbosch, Tygerberg Campus, South Africa, did the mass spectrometric analysis of all the protein samples. They prepared the solutions and buffers needed to perform the analysis.

2.13.2.1 Tryptic digestion of protein

Selected protein spots were cut from the gel manually using a yellow pipette tip or by an automated spot cutter (EXQuest^{TM} Spot Cutter) (BIO-RAD). Gels plugs were put in sterile 1.5 ml Eppendorf tubes. Proteins were reduced with 10 mM DTT for 1 h at 57°C. This was followed by brief washing steps of ammonium bicarbonate followed by 50% acetonitrile before proteins were alkylated with 55 mM iodoacetamide for 1 h in the dark.

Following alkylation, the gel pieces were washed with ammonium bicarbonate for 10 min followed by 50% acetonitrile for 20 min, before being dried in vacuum. The gel pieces were digested with 20 µl of a 10 ng/µl trypsin solution at 37°C overnight. The resulting peptides were extracted twice with 70% acetonitrile in 0.1% formic acid for 30 min, and then dried and stored at -20°C. All peptides were cleaned using stage tips before injection. Dried peptides were dissolved in 5% acetonitrile in 0.1% formic acid and 10 µl injections were made for nano-LC chromatography.

2.13.2.2 Mass spectrometry analysis of protein samples

All experiments were performed on a Thermo Scientific EASY-nLC II connected to a LTQ Orbitrap Velos mass spectrometer (Thermo Scientific, Bremen, Germany) equipped with a nano-electrospray source. For liquid chromatography, separation was performed on a EASY-Column (2 cm, ID 100µm, 5 µm, C18) pre-column followed by a EASY-column (10 cm, ID 75 µm, 3 µm, C18) column with a flow rate of 300 nl/min. The gradient used was from 5-40% solvent B in 20 min, 40-80% solvent B in 5 min and kept at 80% solvent B for 10 min. Solvent A was 100% water in 0.1% formic acid, and solvent B was 100% acetonitrile in 0.1% formic acid.

The mass spectrometer was operated in data-dependent mode to automatically switch between Orbitrap-MS and LTQ-MS/MS acquisition. Data were acquired using the Xcaliber software package. The precursor ion scan MS spectra (m/z 400 – 2000) were acquired in the Orbitrap with resolution R = 60000 with the number of accumulated ions being 1 x 106. The 20 most intense ions were isolated and fragmented in linear ion trap

(number of accumulated ions 1.5 x 104) using collision-induced dissociation (CID). The lock mass option (polydimethylcyclosiloxane; m/z 445.120025) enabled accurate mass measurement in both the MS and MS/MS modes. In data-dependent LC-MS/MS experiments, dynamic exclusion was used with 60 seconds exclusion duration. Mass spectrometry conditions were 1.5 kV, capillary temperature of 200 °C, with no sheath and auxiliary gas flow. The ion selection threshold was 500 counts for MS/MS and an activation Q-value of 0.25 and activation time of 10 ms were also applied for MS/MS.

2.13.2.3 Data analysis

Thermo Proteome Discoverer 1.3 (Thermo Scientific, Bremen, Germany) was used to identify proteins via automated database searching (Mascot, Matrix Science, London, UK) of all tandem mass spectra against the apple peptide database obtained from http://genomics.research.iasma.it/ and a fungal database. Carbamidomethyl cysteine was set as fixed modification, and oxidized methionine, N-acetylation and deamidation (NQ) was used as variable modifications. The precursor mass tolerance was set to 10 ppm, and fragment mass tolerance set to 0.8 Da. Two missed tryptic cleavages were allowed. Proteins were considered positively identified when they were identified with at least 2 tryptic peptides per proteins, a MASCOT score of more than $p < 0.05$ as determined by Proteome Discoverer 1.3. Percolator was also used for validation of search results. In Percolator a decoy database was searched with a FDR (strict) of 0.02 and FDR (relaxed) of 0.05 with validation based on the q-value.

2.13.2.4 Data bases

Apple protein sequences were downloaded from the GDR website and *Venturia inaequalis* protein sequences were obtained from the first draft sequence generated at the University of the Western Cape, Bellville, South Africa. Databases were subsequently generated using the Mascot software (www.matrixscience.com) and sent to Central Analytical Facility (CAF), Proteomics laboratory, at the University of Stellenbosch, Tygerberg Campus, South Africa.

2.13.2.5 Bioinformatic analysis

Theoretical Mr and pI of MS identified proteins were estimated using the Compute pI/MW tool available on ExPASy (www.expasy.org). Proteins were grouped into functional categories (Bevan *et al*., 1998) using data available on the UniProt database (www.uniprot.org) as well as literature sources.

2.14 RNA-Seq analysis of *Venturia inaequalis* infected and uninfected apple leaves

2.14.1 Preparation of libraries

cDNA libraries were prepared following the manufacturer's sample preparation guide (Illumina, Inc, San Diego, California). This protocol involved a number of steps *viz.* purification and fragmentation of mRNA, synthesis of the first cDNA strand, synthesis of the second cDNA strand, end repair, adenylate 3′ ends, ligate adapters, enrich DNA

fragments and validation of libraries. In the first step, poly-T oligo-attached magnetic beads are used to purify the poly-A containing mRNA molecules from the total RNA extracted according the protocol (section 2.8.2). Two rounds of purification were done, with the second elution of the poly-A RNA, the RNA was also fragmented and subsequently primed for cDNA synthesis. In the synthesis of the first strand, reverse transcriptase and random primers were used to reverse transcribe the cleaved RNA fragments into first strand cDNA. Synthesis of the second cDNA strand involved 2 steps *viz.* the removal of the RNA template and synthesis of a replacement strand to generate double-stranded (ds) cDNA. This mixture of ds cDNA and second strand mix was separated with Ampure XP beads. In the end repair process, the overhangs resulting from fragmentation were converted into blunt ends, using an End Repair mix. The 3′ and 5′ exonuclease activity of this mix removed the 3′ overhang and the polymerase activity fill in the 5′ overhangs. The adenylation step of the 3′ end involved the addition of a single adenine (A) nucleotide to the 3′ ends of the blunt fragments to prevent ligation of fragments to one another during the adapter ligation reaction. A corresponding single thiamine (T) nucleotide on the 3′ end of the adapter provided a complementary overhang for ligating the adapter to the fragment. This was followed by a selective enrichment process in which DNA fragments containing the adapter molecules on both ends were amplified and furthermore amplifying the amount of DNA in the library. Finally the cDNA library was validated through agarose gel electrophoresis. Transcriptome sequencing was then performed on the Illumina HiScan SQ machine and MiSeq based at Onderstepoort Veterinary Institute, Agricultural Research Council, Biotechnology Platform, Pretoria.

2.14.2 Tools used in Bioinformatic analysis of data

2.14.2.1 Trimming and cleaning of data

Transcriptomic data received was analysed by first doing a quality check using FastQC version 0.8.0 (Babraham Bioinformatics, 2012). Subsequent data analysis was done using a pipeline of bioinformatic tools. Reads with a PHRED score below 20, adapter sequences, the first 10- and last 5-10 bases were removed using Trimmomatic software (Released version 0.30, http://www.usadellab.org/cms/?page=trimmomatic). A minimum read length of 30 bases on both sides of the paired-end and single end format were subjected for further analysis. Again a quality check was performed using FastQC version 0.8.0.

2.14.2.2 mRNA-Seq read mapping and transcript abundance estimation

In order to estimate transcript abundance the following pipeline was used. The assembled *Malus* x *domestica* genome sequence (Velasco *et al.*, 2010), i.e. genome contig FASTA file and assembly gff3 downloaded from the website http://genomics.research.iasma.it/, was used to estimate transcript abundances. All high-quality reads of each sample were mapped using TopHat2 (Kim *et al.*, 2013) to the *Malus* x *domestic* v.1 reference genome (Velasco *et al.*, 2010). HTSeq (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html), a Python package that provides infrastructure to process data from high throughput sequencing experiments, was used to

‘‘count’’ the number of reads uniquely mapping to each gene. Differential gene expression was inferred based on the total mapping counts using the limma (version 3.17.26) (Smyth, 2004) and edgeR (version 3.3.8) packages (Robinson *et al.*, 2010), implemented in R (version 3.0.2) (R Development Core Team, 2011) taking into account a batch effect between the condition and time point. For some of the data sets, technical replicates were included in the analysis. The technical repeat samples’ counts were added up to get a single column, which correspond to a unique sample. The library sizes in the count table obtained were scaled by edgeR normalisation factors (Robinson *et al.*, 2013). Furthermore, limma/voom from the limma package was used to calculate the mean-variance trend using lowess regression (Smyth, 2004). This function converted the read counts to $\log_2$-cpm (count-per-million) with associated weights ready for linear modelling (Smyth, 2004). Genes were considered differentially expressed (DE) if they possessed an absolute value of $\log_2$-fold change ≥ 1, ≤ 1 and an adjusted p-value ≤ 0.05 (with application of the Benjamini and Hochberg correction for multiple testing), (Benjamini & Hochberg, 2000).

2.14.2.3 Extraction of differentially expressed genes

The following commands and in-house python scripts were used to extract differentially expressed genes from the apple genome. The first script *viz.* convert.py (Appendix B) was used to rearrange the apple gff3 file.

In the subsequent file, mRNA replaced the mRNA2. In the resulting file, the following command:

gffread apple_genome.cleaned.gff3 -F -O -M -K -Q -T -o apple_collapsed.gtf,

was used to collapse the transcripts and exons into genes. However, in the apple genome gff3 file were only the loci (RLOC). The alignment using Tophat v 2 gave as output bam files, which could be used to generate a count table and was subsequently used as input in limma/edgeR. The following commands were used: (i) the bam files were sorted

samtools sort accepted_hits.bam accepted_hits_sorted.bam

followed by indexing:

samtools index accepted_hits_sorted accepted_hits_sorted.bam.bai

The following command was used to generate the count_table:

samtools view C0_tophat_out/accepted_hits_sorted.bam | htseq-count -m intersection-strict -s no -i locus - apple_genome.gtf > counts_C0.

This command was used for each data set i.e. C0, E0, C2, E2, C4, E4, C8, E8, C12, E12. The output from this *viz.* count_0 to count_12 was merge using the merge_count.py script (Appendix B).

The resultant file was used in limma. The commands in Appendix C were used in limma/voom and edgeR to generate a MDS plot, heat map, Venn diagram and a list of DE genes.

2.14.2.4 Scripts used in analysing the Data

In-house perl and python scripts were used to analyse the data (See Appendix B for scripts).

2.14.2.5 Identification of differentially expressed genes

A list of 398 differentially expressed genes due to condition was generated by the limma analysis. The corresponding loci were extracted from the reference genome (http://genomics.research.iasma.it/). Refer to Appendix B for a step-by-step workflow of scripts used. The perl script used extracted a list of candidate differentially expressed genes from the apple genome. These differentially expressed genes were then submitted to Blast2GO (Conesa and Gotz, 2008).

2.14.2.6 Functional annotation

Functional annotation for all host-pathogen interaction genes were generated using the Blast2GO suite (http://www.BLAST2GO.org/), Bioinformatics and Genomics Department, Centro de Investigación Príncipe Valencia, Spain; (Conesa and Gotz, 2008; Conesa *et al.*, 2005). The annotation procedure consists of three main steps i.e. (i) blast to find homologous sequences, (ii) mapping to collect gene ontology (GO) terms associated to blast hits and (iii) annotation to assign trustworthy information to query sequences. In the Blast step the non-redundant (nr) nucleotide database was selected, blast was performed remotely on the NCBI server through Qblast. Blast parameters were kept at default values: *e*-value threshold of 1*e*-3 and a recovery of 20 hits per sequence. The minimal alignment length (hsp) filter was set at 33 to avoid hits where the length of the matching region is smaller than 100 nucleotides. A mapping step was performed in which GO terms were retrieved for the hits obtained after blast search.

Following mapping, an annotation step was performed with *e*-value filter at 1*e*-3, default gradual evidence code (EC) weights, a GO weight of 15 with an annotation cutoff of 60. Furthermore, InterPro Scan (Berardini *et al.*, 2004) was performed to find functional motifs and related GO terms for the Blast-based GO annotations by using the specific tool implemented in the Blast2GO software with the default parameters. These results were merged to the already existing annotation and completed with an implicit annotation done through the ANNEX function http://goat.no, (Zdobnov & Apweiler, 2001). The GOslim 'goslim_plant.obo' was used to achieve specific GO terms by means of a plant-specific reduced version of the Gene Ontology http://geneontology.org.

2.14.2.7 Enzyme mapping

Enzyme mapping of annotated sequences was done by direct GO to Enzyme annotation and used to query the Kyoto Encyclopaedia of Genes and Genomes (KEGG - http://www.genome.jp, (Kanehisa & Goto, 2000; Kanehisa *et al.*, 2006; Kanehisa *et al.*, 2008) to define the main metabolic pathways involved.

Chapter Three - Genotyping of a *Venturia inaequalis* isolate using 21 polymorphic microsatellite markers

3 Introduction

Venturia inaequalis (Cke) Wint, a heterothallic, haploid ascomycete, is the causal agent of apple scab, an important disease, which occurs in every country where apples are produced (Machardy, 1996). This disease could cause huge economic losses of up to 70% reduction in apple production (Biggs, 1997). Scab infection causes fruit to be deformed in shape and size of the fruits, premature leaf and fruit fall, leaving trees vulnerable to chilling and freezing injuries (Jha *et al.* 2010). The pathogen has been categorized into eight physiological races (section 1.4, Table 1.3) based on its hypervariability and the exhibition of differential pathogenicity on apple cultivars (Jha *et al.* 2010). Some races are capable of growing on two different differential hosts making it difficult to classify them to a particular race.

Molecular tools such as RADP, PCR-RFLP, AFLP and microsatellites have been used to reveal genetic diversity amongst isolates collected from different regions of the world. Microsatellites, also known as simple sequence repeats (SSRs) are tandemly repeated nucleotide motifs embedded in unique DNA stretches (Tautz, 1989). Their hypervariability, abundance in eukaryotic genomes, and co-dominant nature, make SSR markers a powerful tool for taxonomic, population genetic studies and genotyping of an isolate. Guérin *et al.* (2004) developed and published polymorphic SSR markers specific

for *Venturia inaequalis* and were successfully used in population genetic studies to demonstrate population diversity (Gladieux *et al*., 2008; Gladieux *et al*., 2010). The objective of this study was, therefore, to verify that the isolate used in the study, was *Venturia inaequalis*, by using SSR markers.

3.1 Establishment of fungal culture

In order to pursue the objectives set in the study, the isolate had to be purified and verified/characterised. For this purpose a fungal culture was established (section 2.6) to extract genomic DNA (section 2.6.1) from the mycelia. Fungal spores were isolated as depicted in Fig. 3.1. These single isolated spores were placed on water agar medium and allowed to germinate. Germinated spores were placed on potato dextrose agar (PDA) medium plates and allowed to grow. Gel plugs containing actively growing mycelia were cut from the PDA plates and placed in potato dextrose broth, allowed to grow while agitating for 2-3 months as depicted in Fig. 3.2. Genomic DNA was extracted from mycelia as described in section 2.6.1 and the sample was genotyped using 21 microsatellite loci (Guerin *et al.*, 2004) (Table 3.1).

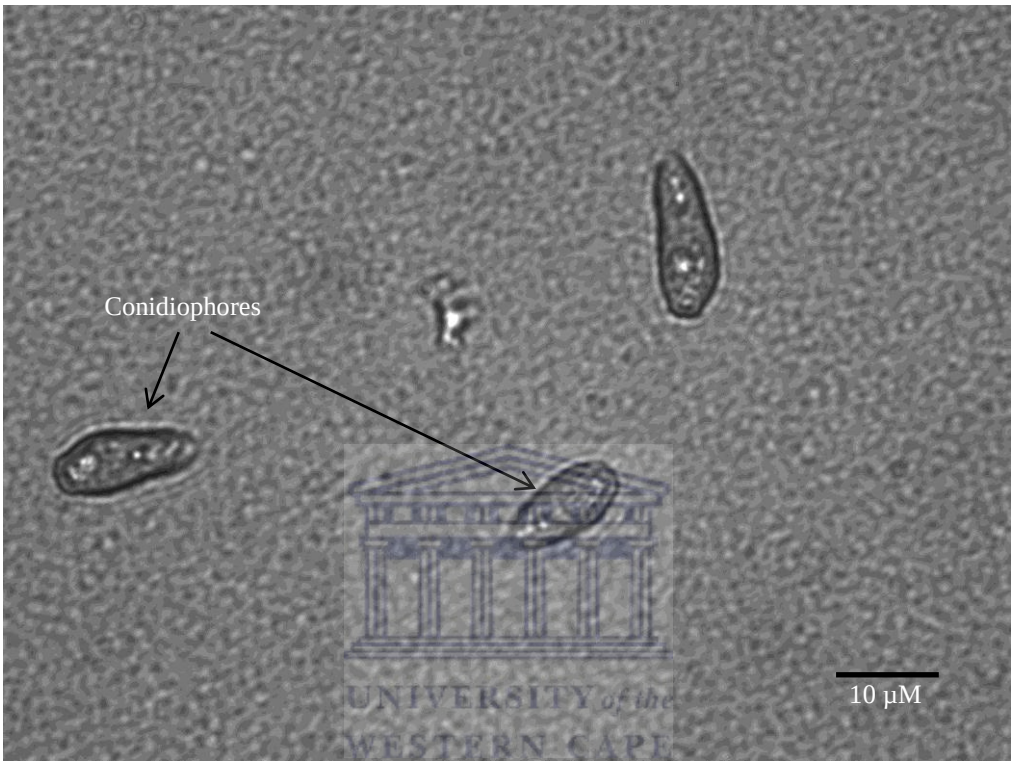


Figure 3.1 Typical flame shaped conidiophores of *Venturia inaequalis* as observed under the microscope.

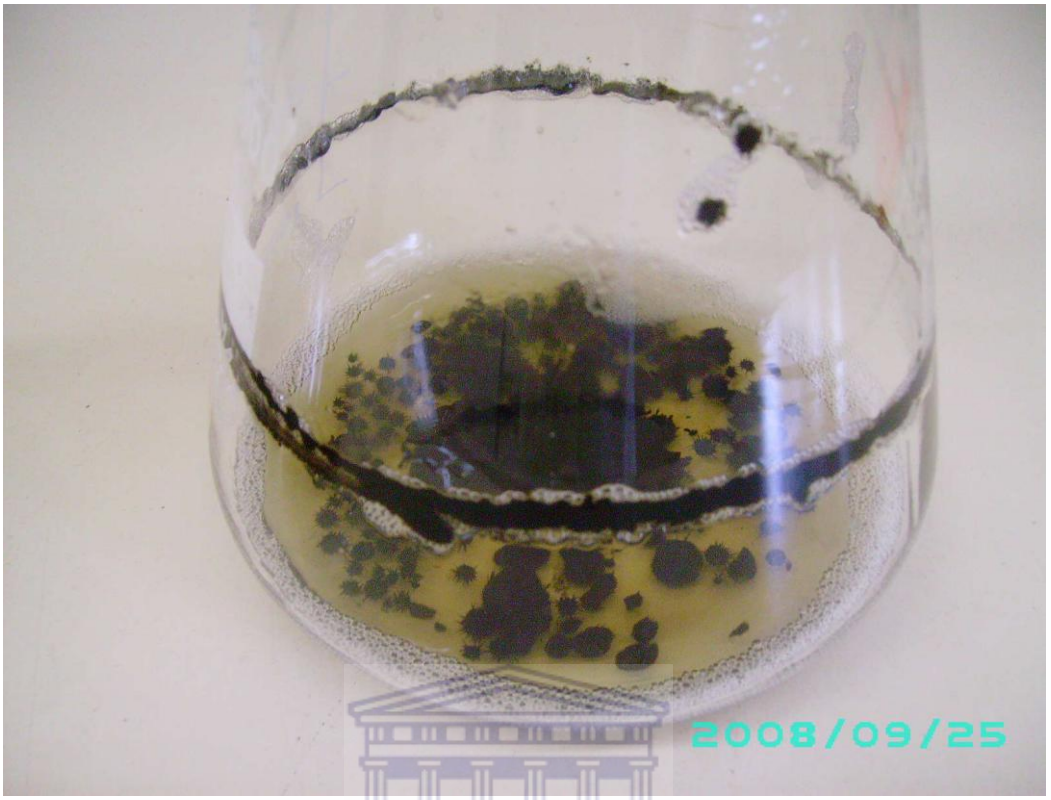


Figure 3.2 *Venturia inaequalis* fungal culture. Flask containing clumps of mycelia growing in potato dextrose broth.

3.2 Polymorphic markers and electropherograms

Alleles were amplified for all the markers within the specified ranges for the markers (Table 3.1 and Fig. 3.3). Electropherograms have the advantage that amplified alleles, which differ in size of 1 bp can be seen (Fig. 3.3).

Table 3.1 Published 21 polymorphic microsatellite markers (Guerin *et al.*, 2004) used to characterise the WS-strain.

Primer Name	*Locus*	*Size Range (bp)*	*Observed Size (bp)*	*Within size Range*
1	*Vitc1/2*	173-241	191	X
2	*Vitc1/82*	227-247	232; 233	XX
3	*Vitc1/130*	132-152	150	X
4	*Vitc2/D*	184-278	219	X
5	*Viaggt8/1*	188-196	181; 196	XX
6	*Vica9/152*	167-191	170	X
7	*Vitcca7/P*	192-224	191	X
8	*Vitg11/70*	184-196	185; 200	XX
9	*Vica9/134*	228-236	234	X
10	*Vigtg10/95*	134-169	124; 135	XX
11	*Vicacg8/42*	196-232	200	X
12	*Vigt8/146*	128-134	132	X
13	*Vitc2/16*	147-165	135	X
14	*Viga3/Z*	87-97	95	X
15	*Viga7/116*	159-173	152	X
16	*Vitg9/99*	155-167	160	X
17	*Vica9/X*	225-239	230	X
18	*Vica10/154*	108-172	178	X
19	*ViaacS10*	180-186	168; 181	XX
20	*Vigt10/ε*	171-173	173	X
21	*Vitg9/129*	267-285	256; 276	XX

X - denotes one allele and XX - denotes two alleles amplified

Table 3.2 Comparative table of SSR loci amplified in European strains (Broggini *et al.*, 2010) versus South African strain (WS-Strain)

SSR locus	*Label*	*Size Range (bp)*	*Alleles (bp)*			*Linkage Group*
			1639	EU-BO4	WS-Strain	
Viaggt8/1	FAM	188-196	197	195	181; 196	LG-5
Vica10/154	HEX	108-172	143	172	178	Vi-4
Vica9/152	HEX	167-191	177	170	170	LG-5
Vicacg8/42	FAM	196-232	210	216	205	LG-5
Vitc1/130	FAM	132-152	154	158	150	LG-5
Viga7/116	FAM	159-173	163	161	152	LG-5
Vigtg10/95	FAM	134-169	135	143	124; 135	LG-12
Vitc1/82	HEX	227-247	238	236	232; 233	Vi-4
Vitc2/D	HEX	184-278	Absent	228	219	Vi-6
Vitcca7P	FAM	192-224	194	Absent	191	Vi-4
Vitg11/70	HEX	184-196	187	191	185; 200	Vi-11
Vitg9/129	FAM	267-285	278	280	276	LG-5
Viga3/Z	FAM	87-97	95	95	95; 454; 471	-
Vica9/134	HEX	228-236	231	231	234	-
Vica9/X	HEX	225-239	234	234	230	-
Vitg9/99	HEX	155-167	158	158	160	-

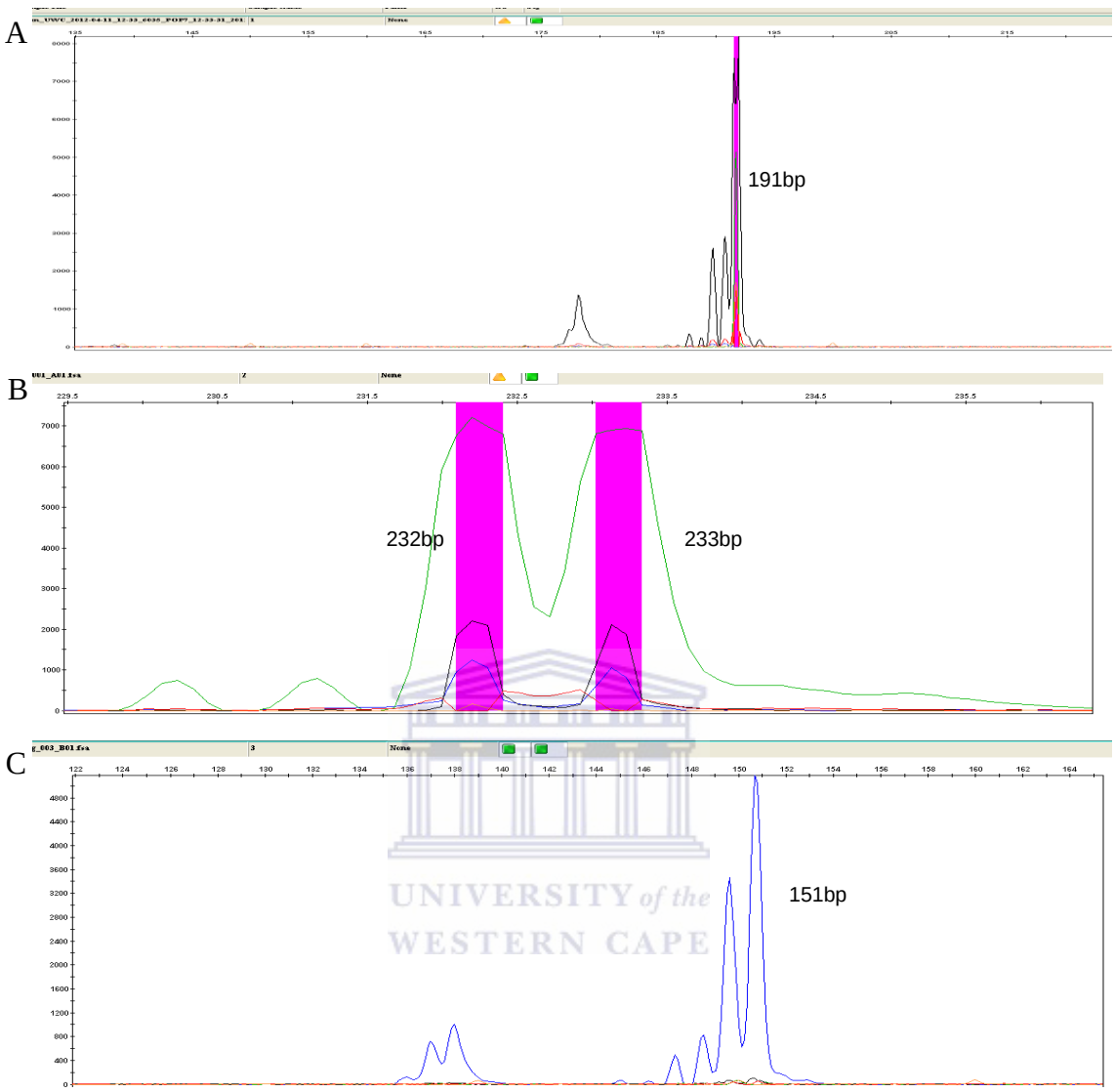


Figure 3.3 Electropherograms depicting alleles amplified. (A) Shows one allele amplified with marker *Vitc1/2*. (B) Shows two alleles amplified with marker *Vitc1/82*. (C) Shows one allele amplified with marker *Vitc1/130*. See Table 3.1 for the remainder of markers amplified.

3.3 Discussion

3.3.1 Establishing of *in vitro* culture

Culturing *Venturia inaequalis* stand the chance of being contaminated by other fungi, therefore purification and verification of the strain was of utmost importance for the isolation of gDNA. A characteristic of *Venturia inaequalis* is that its a very slow growing fungus, and takes 5-10 days for growth. Other fungi like *Alternaria* for example are very fast growing and spread within 3-4 days over the agar plate. Detailed investigations were conducted on the morphological features of the fungus. The spores had a characteristic flame shape, in all the isolates (Figure 3.1). However, the spore size was variable with sizes ranging from 13–36 x 6–12 μM.

3.3.2 Molecular characterisation of *Venturia inaequalis*

Characterisation of the *Venturia inaequalis* strain was crucial for the entire project since the credibility of results depended on the purity of the strain/isolated investigated. For this exercise, published microsatellites Guerin *et al*. (2004) were used. All of the primer pairs amplified loci, which fell within the size ranges of the specified primer. However, six primer sets showed polymorphism or artefacts.

In a study performed by Broggini *et al.* (2010), sixteen of the 21 published polymorphic microsatellite markers (Guerin *et al.*, 2004) used in their study were found to have alleles in the EU-BO4 and 1639 isolates in Table 3.2 (Broggini *et al.*, 2010). Some of these alleles mapped onto some of the 15 major linkage groups as mentioned in Broginni *et al.*

(2010). The same 16 corresponding markers were found to have alleles of the isolate in question and mapped on the same linkage groups namely LG 5 and LG 12 as in the Broggini *et al.* (2010) study.

Gladieux and his group (2008) have used 6 of the published microsatellite markers (Table 3.2) (Guerin *et al.,* 2004) to cluster the samples into regions of origin (Gladieux *et al.*, 2008). In their findings they showed that the South African samples clustered in the same region as the European haplotypes (Gladieux *et al.*, 2008). Thus the strains used in the Broggini study and this study would than cluster in the same cluster, based on these findings (Table 3.2) and on those of Gladieux *et al.* (2008).

In a study performed by Celton *et al.* (2010), a draft genome sequence of the haploid *Venturia inaequalis* isolate was used to develop a SNP scoring strategy. In this strategy, the draft genome served as a reference genome in anchoring contigs to SNPs and the validation of the strategy was performed on a *Venturia inaequalis* bin population (Celton *et al.*, 2010). Furthermore, the *Venturia inaequalis* genetic used in the Celton *et al.* (2010) study, was used for this study as well.

In conclusion, amplification of all the expected alleles (Table 3.1) is confirmation that the isolate is a *Venturia inaequalis* isolate. Results obtained from Broggini *et al.* (2010) further support this finding by mapping the alleles in the same linkage groups as theirs. Furthermore, Gladieux *et al.* (2008) also support these results by placing the South African strains used in their study within the European cluster. The strain used in this

study also amplified alleles, which fell within the range as those of the European strain. Work by the Celton group (2010) also add additional confirmation for the isolate used to be *Venturia inaequalis*.

Chapter Four - A transcriptomic analysis of the interaction between *Venturia inaequalis* and susceptible *Malus* x *domestica* (Borkh) cv Golden Delicious

4 Introduction

The fungal pathogen *Venturia inaequalis* establishes a biotrophic interaction with its host plants, including apple (*Malus* x *domestica*). It causes scab lesions on leaves, deformation in shape and size of fruit, premature leaf and fruit fall with subsequent susceptibility to chilling, which results in serious economic losses. Global expression profiling using RNA-Seq and the recently sequenced apple genome identified a complex network of genes involved in the apple-*Venturia inaequalis* interaction. Messenger RNA from apple leaves infected with *Venturia inaequalis*, sampled at five time points post inoculation were used for the identification of differentially expressed genes putatively involved in disease resistance. cDNA libraries were constructed for control and treated plants and sequenced using RNA-Seq technology. The RNA-Seq data was analysed for differentially expressed genes using TopHat2, edgeR and limma/voom. Eighty six percent of the up-regulated and 82.2% down-regulated genes were similar to known proteins in the non-redundant database. The observed high number of differentially expressed genes allowed us to classify them according to the biological pathways in which they are implicated, generating a holistic picture. BLASTx analysis, GO terms mapping and annotation analysis were performed using the Blast2GO software, a research tool designed with the main purpose of enabling GO based data mining on

sequence sets for which no GO annotation is yet available. Most pronounced was the up-regulated genes, which were highly expressed could be associated with phenylalanine metabolism, phenylpropanoid metabolism, starch and sucrose metabolism. A closer look at which genes were differentially expressed, revealed genes involved in signal perception, transcription, detoxification, defence, resistance, transporters, structural modifications and secondary metabolites.

4.1 Inoculation and collection of seedlings

Young Golden Delicious apple seedlings were inoculated with *Venturia inaequalis* strain. It were sprayed using a low-pressure gun in an infection chamber with the control plants separate from the infected plants, as discussed in section 2.5. Seedlings for control were mock inoculated with sterile water. Apple leaves for different time points starting 0 to 12 days post inoculation (dpi) were collected in liquid nitrogen as discussed in section 2.5.2.

4.2 Isolation of total RNA from *Venturia inaequalis* infected and uninfected leaves

The youngest leaves from the seedlings were collected and immediately frozen in liquid nitrogen. Total RNA was isolated following a modified protocol of Menhaj *et al*. (1999) as discussed in section 2.8.2. The RNA was DNase treated to degrade genomic DNA present in the extraction as discussed in section 2.8.2.

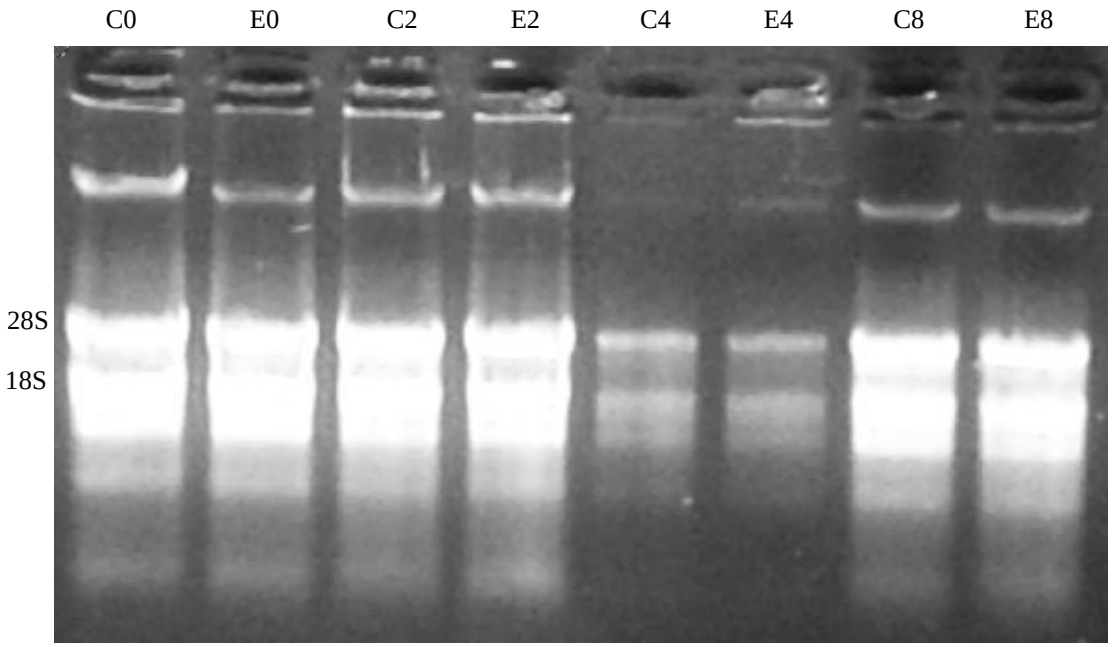


Figure 4.1 Total RNA isolated from apple leaves (infected and uninfected) with *Venturia inaequalis*. Gel depicts total RNA extracted from control and infected apple leaves.

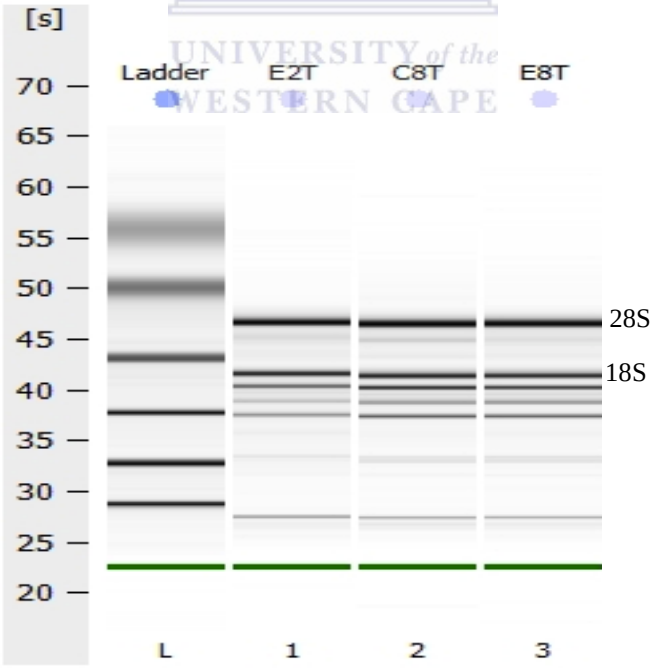


Figure 4.2 Bioanalyzer image of total RNA isolated from infected and uninfected apple leaves. Lanes L shows the profile of the RNA molecular ladder, 1 shows sample E2, 2 is sample C8 and 3 is sample E8.

Total RNA isolated was of good quality with the 28S and 18S showing intense bands. In the agarose gel, genomic DNA is present in the RNA. The RNA was treated with DNAse as can be seen in the Bioanalyzer image with gDNA absent. The quality of the RNA is good since the 28S and 18S bands appeared at the 46 seconds for 28S and 43 seconds for 18S (Fig. 4.2) respectively.

4.3 mRNA-Seq data analyses

Expression profiling of *Venturia inaequalis* infected and uninfected apple leaves at five time points (Day 0, 2, 4, 8 and 12) were performed with different sequencing instruments namely, Illumina HiScan SQ™ and MiSeq™. This resulted in a mixture of single-end and paired-end reads respectively for the analysis. The total number of reads for each time point ranged from 2 to 21 million reads. A total of 128 056 666 high-quality reads (average length = 90 bp) for control (untreated) and 143 292 730 high-quality reads (average length = 90 bp) (Table 4.1) for infected (treated) leaf tissue were generated using these sequencers. These reads were mapped against the apple genome (Velasco *et al*., 2010) which served as a reference genome.

4.4 Data quality control (QC) assessment

Single and paired end reads generated from the HiScan SQ™ and the MiSeq™ instruments are tabulated in Table 4.1. The raw reads of each data set were trimmed for Illumina adapters and subsequent low quality bases were filtered out using Trimmomatic version 0.30. The quality trimmed reads ($Q \geq 20$) were then aligned to the complete apple genome using Tophat2 (Kim *et al.*, 2013) as discussed in 2.14.2.1.

Table 4.1 Summary of read mapping for single-end and paired-end reads against the apple genome sequence

Time point and treatment	*Experiment*	*Total high-quality reads generated*	*Total reads per biological condition*	*Total reads mapped to apple reference*	*%*	*Overall reads mapped*
Day 0	Untreated	SE-14108823		13620228	96.5	96.5
		fwd-6834913	27778649	3433571	50.2	50.3 (av.73%)
		rev-6834913		3440581	50.3	
	Treated	SE-14108823	27674213	13620228	96.5	96.5
		fwd-1133896		489060	43.1	43.9
		rev- 1133896		506883	44.7	
		fwd-5648799		3057275	54.1	54.2 (av.64%)
		rev-5648799		3062563	54.2	
Day 2	Untreated	fwd-21821111	43642222	20448487	93.7	93.6
		rev-21821111		20381924	93.4	
	Treated	fwd-768370		620275	80.7	80.0
		rev-768370		607567	79.4	
		fwd-7287652	29748076	3415118	46.9	47.5
		rev-7287652		3514911	48.2	
		fwd-6818016		3916994	57.5	57.4 (av.61%)
		rev-6818016		3907168	57.3	
Day 4	Untreated	SE-18056034	18056034	17583271	97.4	97.4
	Treated	SE-19327833	19327833	18524410	95.8	95.8
Day 8	Untreated	fwd-6726298		5980895	88.9	88.8
		rev-6726298	31228482	5959858	88.6	
		fwd-8887943		3933073	44.3	44.1 (av.66%)
		rev-8887943		3914934	44.0	
	Treated	fwd-3482336		2401339	69.0	68.8
		rev-3482336		2392723	68.7	
		fwd-2318224		693496	35.9	30.7
		rev-2318224	25987980	732139	31.6	
		fwd-7193430		3010444	41.8	41.9 (av.47%)
		rev-7193430		3012683	41.9	
Day 12	Untreated	SE-2220867 fwd-2565206		2329483	90.8	90.3
		rev-2565206	7351279	2301192	89.7	
				1869480	84.2	84.2 (av.87%)
		fwd-20277314	40554628			
	Treated	rev -20277314		10785487	53.2	52.9
				10651999	52.5	
Average Untreated			25611333			92.2%
Treated			28658546		69.2%	60.9%
Total Untreated			128056666			
Treated			143292730			

4.5 Analysis of RNA-Seq data of apple leave after inoculation with *Venturia inaequalis*

In order to identify transcriptional changes associated with *Venturia inaequalis* infection, an expression-profiling experiment was performed using RNA-Seq technology. Analysis of data was performed using a pipeline of bioinformatics tools. These include mapping to the reference genome using Tophat2 (Kim *et al*., 2013), estimation of transcript abundance using edgeR (Robinson *et al*., 2013) and limma/voom (Smyth, 2004), extraction and identification of differentially expressed (DE) genes using in-house perl scripts and functionally annotate the DE genes using Blast2GO® version 2.7.0 (Conesa *et al*., 2005).

4.5.1 Data mapping and transcript abundance estimation

Processed reads as discussed in 4.4 of each time point namely Day 0 to Day 12 for control and treated samples, were mapped against the apple genome as discussed in section 2.14.2.2 using Tophat2 (Kim *et al*., 2013) with the default aligning parameters. Read counts were arranged into a count table using htseq-count, a script supplied by HTSeq, a statistical package (Anders, 2011). This count table was used in subsequent edgeR and limma/voom analysis.

Statistical analysis of the mapped read counts with edgeR (Robinson *et al*., 2013) and limma/voom (Smyth, 2004) revealed a number of differentially expressed genes as a result of the treatment (control versus *Venturia inaequalis* infection), which are depicted

in a heat map that shows how these samples are clustered according to the differentially expressed genes (Fig. 4.2). Samples generally clustered according to either time point following treatment or treatment group. Day 2 (Treatment2) is an apparent exception to this, suggesting a particularly pronounced interaction between the inoculation and time following inoculation.

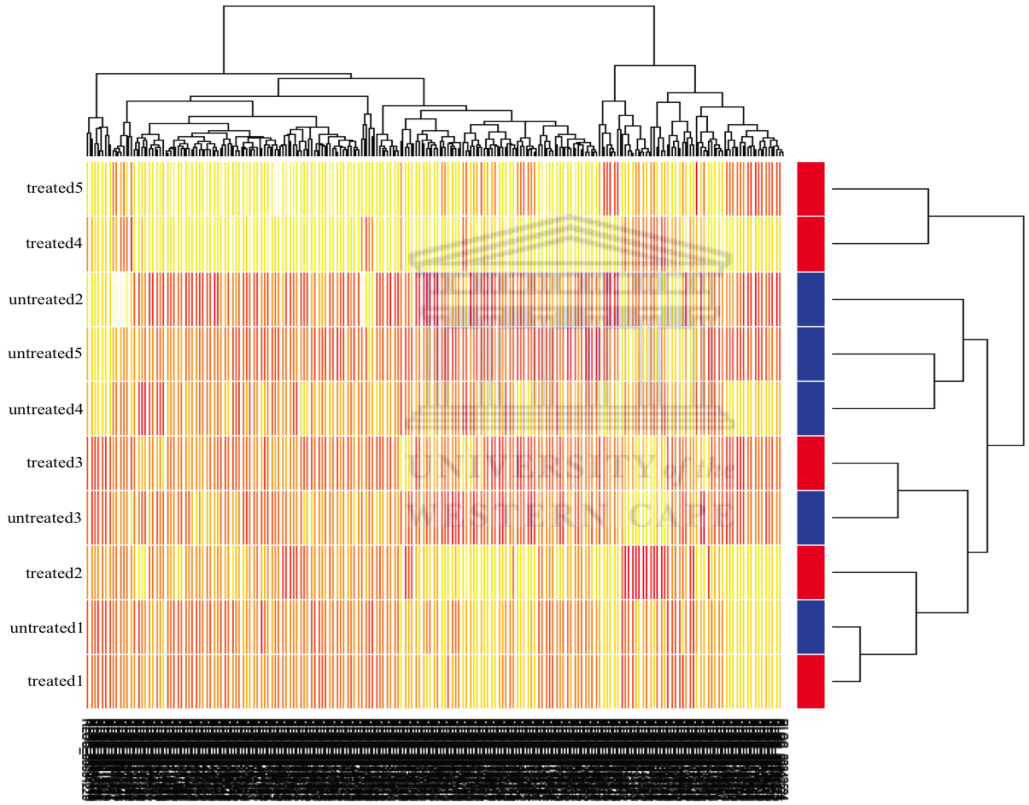


Figure 4.2 Heat map generated in limma/voom for RNA-Seq data. Treated/untreated1 is Day 0, treated/untreated2 is Day 2, treated/untreated3 is Day 4, treated/untreated4 is Day 8, treated/untreated5 is Day 12. Changes of gene expression are displayed from yellow (down-regulated) to red (up-regulated). Lighter colors indicate genes with low fold-change and darker colors indicate genes with high fold-change. Blue color in the cladogram indicates down regulated genes and the red color, up-regulated genes.

The software also allows visualising the data on a multi-dimensional scaling (MDS) plot. A MDS plot (Fig. 4.3) was drawn to visualise the distance between pairs of samples. Distances on the plot correspond to the biological coefficient of variation (BCV) between each pair of samples (edgeR users's guide 2.10). Samples generally clustered according to either time point following treatment or treatment group.

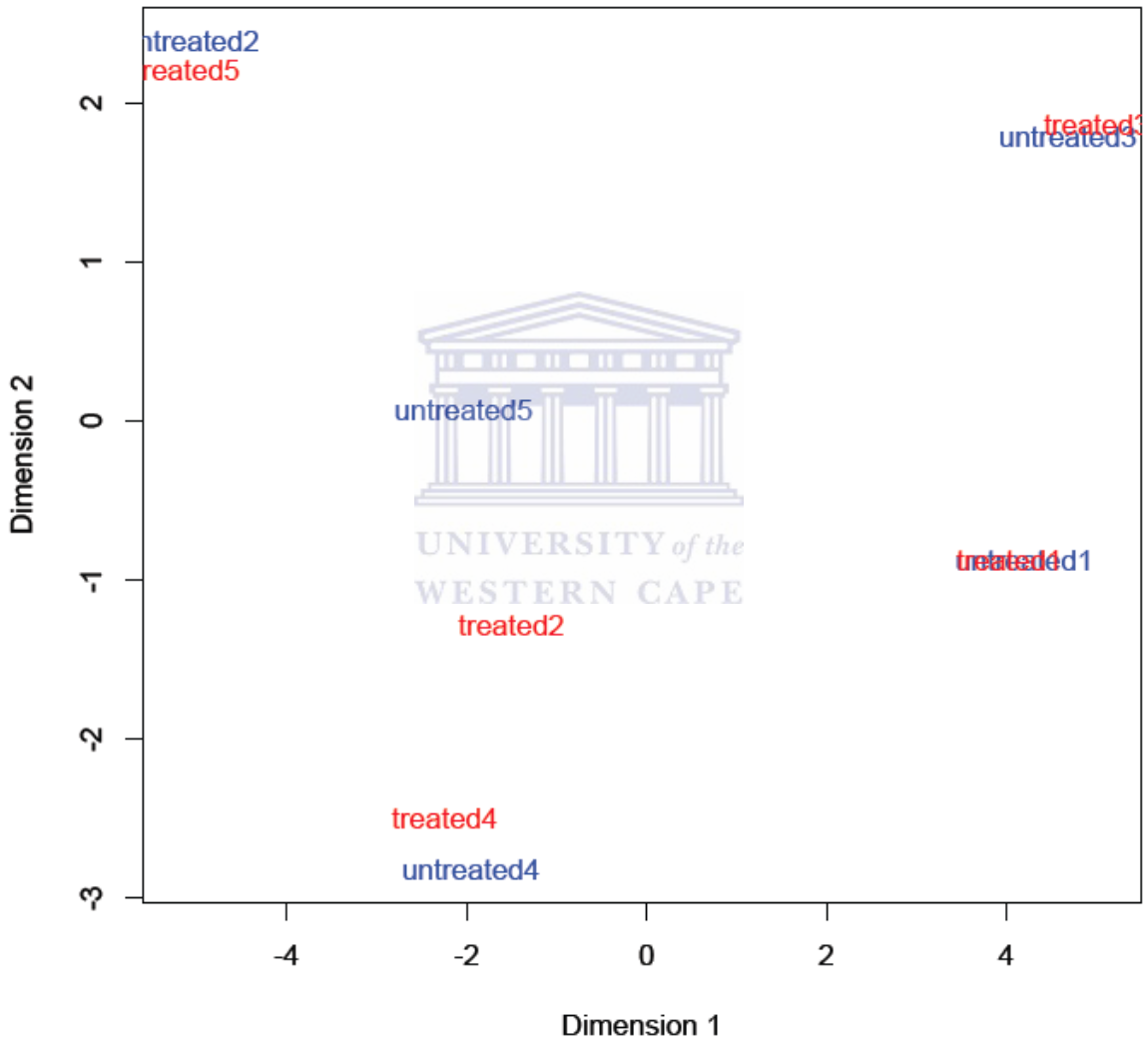


Figure 4.3 Multi-dimensional scaling plot. Samples collected are represented in blue and red color. Blue is the control, red is apple inoculated with *Venturia inaequalis* and 1-5 indicates the days at which samples were collected. Treated/untreated1 represents day 0, treated/untreated2 is day 2, treated/untreated3 is day 4, treated/untreated4 is day 8 and treated/untreated5 is for day 12.

Day 2 (Treatment2) is an exception to this, suggesting a particularly pronounced interaction between the inoculation and time following inoculation then. This also demonstrates that there are a large number of time-dependent changes that must be accounted for.

A Venn diagram (Fig. 4.4) was drawn to depict the number of 9 660 differentially expressed (DE) genes for the time points. Common DE genes are found in the overlapping segments. In this Venn diagram there are DE genes which are uncommon in every time point namely Timepoint 2 (1909), Timepoint 4 (499), Timepoint 8 (1072) and Timepoint 12 (985) of which the highest number of DE genes (1909) is in Timepoint 2. This is expected as the host reacts to the invading fungus. There are 26 DE genes, which are only common in timepoint 2 and 4. A high number of 550 DE genes are only common between Timepoint 2 and Timepoint 12. There are only 15 DE genes, which are common in all the timepoints namely 2, 4, 8 and 12.

In total, 9 660 genes were differentially expressed, of which 5 362 (55.5%) were down-regulated and 4 298 (45.5%) were up-regulated (Table 4.2). Controlling for treatment differences, there were an exceptionally high number of genes differentially expressed at time point 2. At this time point, 1 851 and 1 698 genes were up- and down-regulated versus time point 0, respectively. More genes were generally down-regulated for most of the time points, with the exception of time point 12, in which more genes were up-regulated.

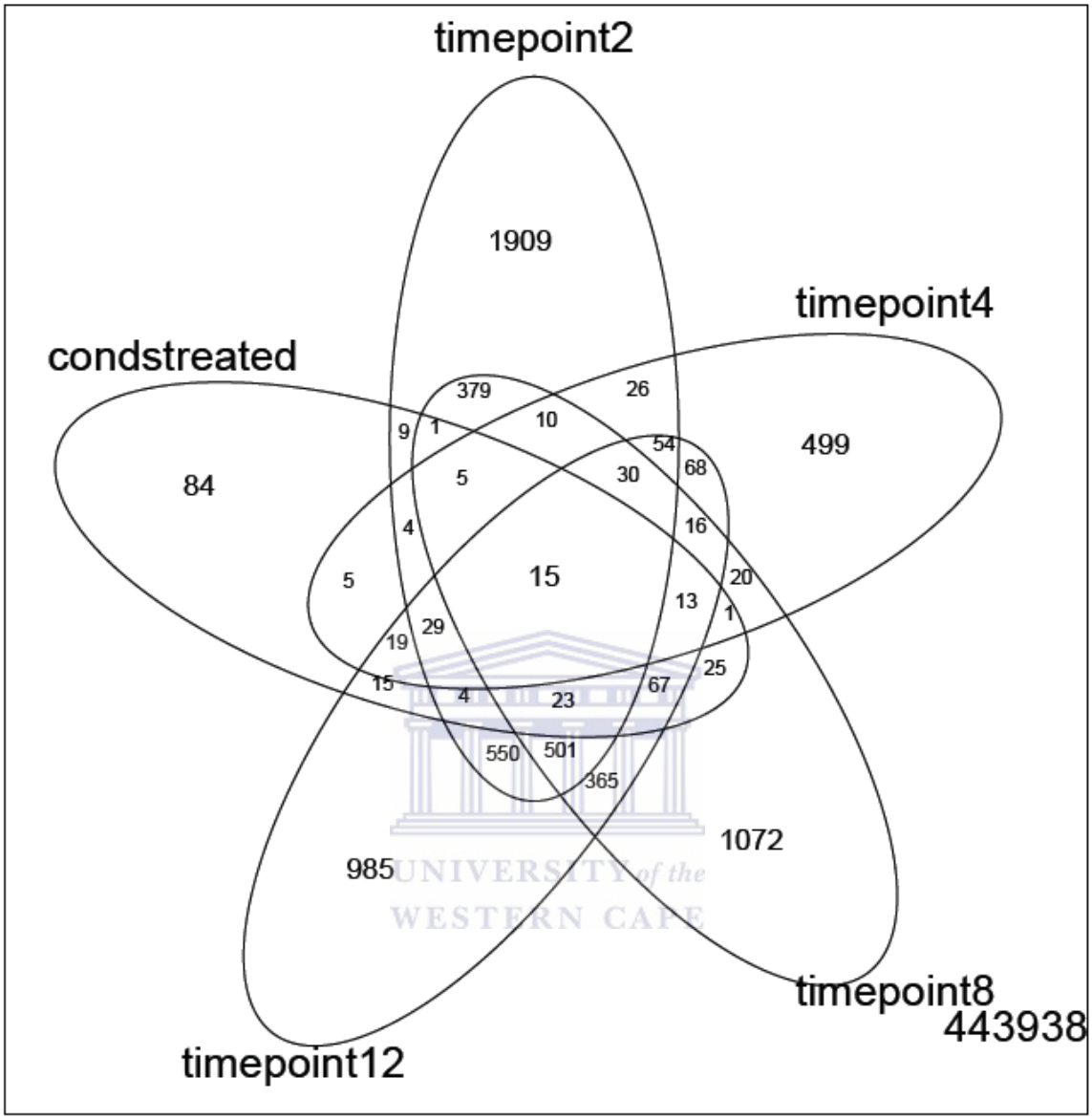


Figure 4.4 Venn diagram depicting putative time point specific and common differentially expressed genes. The number of common genes is shown in the overlapping segment of the Venn diagram. Threshold of adj.P.Val < 0.05 is used.

Table 4.2 Significantly modulated genes (adjusted P-value < 0.05) due to treatment or subsequent time point versus baseline.

	Condstreated	*Timepoint2*	*Timepoint4*	*Timepoint8*	*Timepoint12*
-1	56	1851	652	1661	1198
0	450 422	447 192	449 927	448 198	447 987
1	263	1 698	162	882	1 556

After compensating for the effect of time point, there were 398 genes differentially expressed due to the treatment. Of these genes, 67 were down-regulated and 331 up-regulated. These genes also mapped to 1 164 transcript identities in the apple gene-finding format (gff3) file (http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0), which were submitted to Blast2GO® version 2.7.0 (Conesa *et al*., 2005) for functional annotation.

4.5.2 Extraction and identification DE genes

A number of commands and in-house scripts (Appendix B) were used to extract the differentially expressed genes generated in Table 4.2 (section 4.5.1). Extracted transcripts had to be mapped to the apple genome to extract the corresponding gene from the reference genome Appendix B. These sequences were then submitted to Blast2GO® version 2.7.0 (Conesa *et al*., 2005).

4.6 Functional annotation

The effect of the treatment as discussed in 4.5.1 resulted in 398 differentially expressed (DE) genes of which 331 were up-regulated and 67 were down-regulated. The 331 genes mapped to 1046 transcripts and the 67 genes mapped to 118 transcripts in the apple gene-finding format (gff3) file (http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0). The transcripts obtained from the apple genome as discussed in 4.5.2 were functionally annotated by using Blast2GO® version 2.7.0 (Conesa *et al*., 2005)**,** for BlastX and gene ontology (GO) analysis, as discussed in detail in section 2.14.2.6.

Differentially expressed genes were classified using gene ontology (GO; Gene Ontology Consortium, 2000). Three main categories of GO classification, i.e. biological process, molecular function and cellular component, were analysed separately to understand their functional distribution. GO terms were assigned to the apple genes using Blast2GO version 2.7.0. Of the 1046 transcript sequences submitted, 901 (86.1%) of these genes obtained a BLAST hit with a stringent e-value cut-off ($1e^{-100}$) and to 77.8% of these genes, GO terms were assigned. GO-term enrichment in the set of differentially up- and down- regulated genes were evaluated to indicate which biological processes, molecular functions and cellular components were most affected after inoculation with *Venturia inaequalis*.

The functional distribution of the up regulated genes was further simplified by the assignment of GO-slim terms to the annotated sequences (Harris *et al.*, 2004) of plants.

The up-regulated genes were classed into three gene ontology categories i.e. biological processes, molecular function and cellular component. For the biological process category: metabolic process (GO:0008152), cellular process (GO:0009987), response to stimulus (GO:0050896), single-organism process (GO:0044699) and localisation (GO:0051179) represented the major proportion (Fig. 4.5). Under the molecular function category, catalytic activity (GO:0003824) and binding (GO:0005488) were the top two most abundant subcategories (Fig. 4.6) For the cellular component category, cell (GO:0005623), membrane (GO:0016020) and organelle (GO:0043226) were the most representative level 2 GO terms in the up-regulated genes data set (Fig 4.7). All annotated sequences were then associated with enzyme codes (ECs), which returned 442 unique EC numbers.

To further identify the biological pathways that are active during fungal infection, genes sets were searched against pathway collections in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. A total of 65 pathways were retrieved using Blast2GO version 2.7.0. The most representative pathways included phenylalanine metabolism, phenylpropanoid metabolism, starch and sucrose metabolism, carbon fixation in prokaryotes, propanoate metabolism, purine metabolism, aminobenzoate degradation, lysine biosynthesis, oxidative phosphorylation, riboflavin metabolism and flavonoid biosynthesis (Table 4.3 – Appendix A).

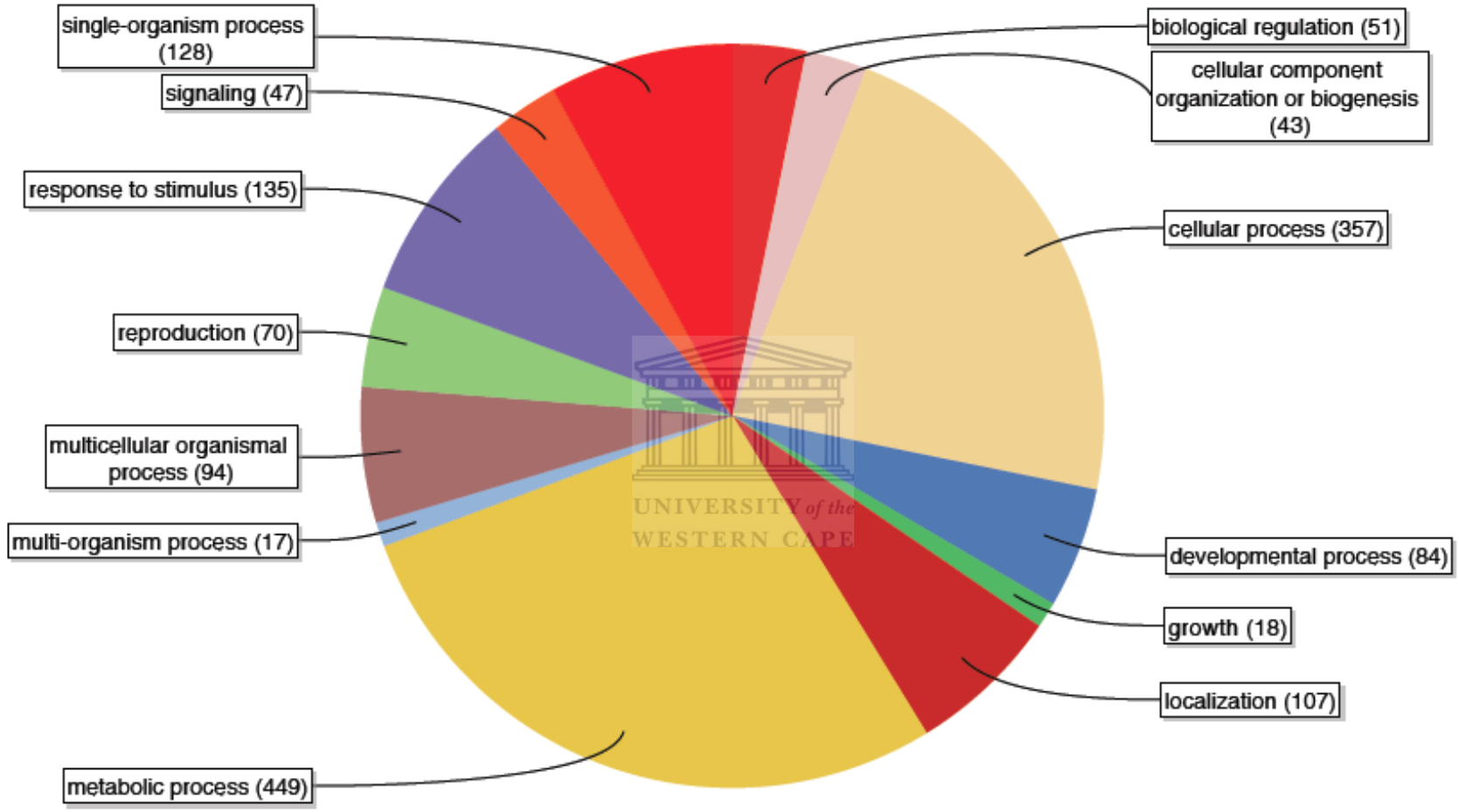


Figure 4.5 GO terms associated with the up-regulated transcripts. Biological processes level 2 up-regulated upon infection with *Venturia inaequalis*.

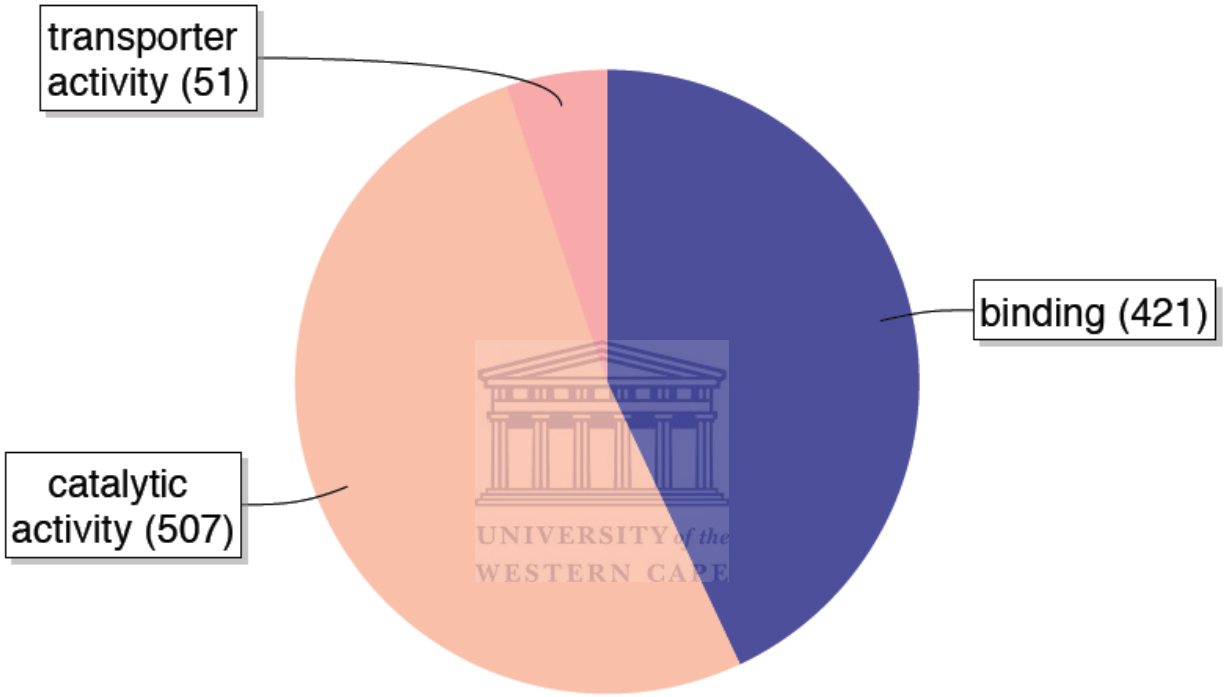


Figure 4.6 GO terms associated with the up-regulated transcripts. Molecular function level 2 up-regulated upon infection with *Venturia inaequalis*.

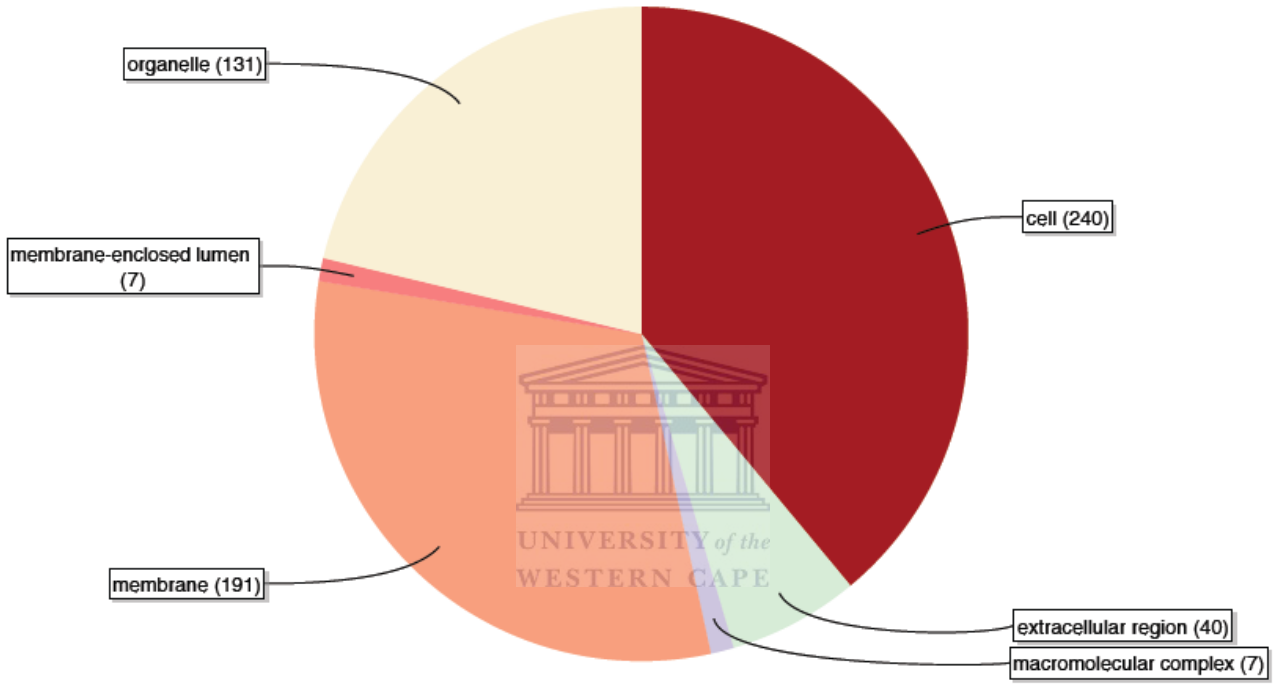


Figure 4.7 GO terms associated with the up-regulated transcripts. Cellular components level 2 up-regulated upon infection with *Venturia inaequalis*.

For the down-regulated genes, annotation was followed as discussed in 4.6. For the biological processes category the following were affected: metabolic process (GO:0008152) and cellular process (GO:0009987) (Fig 4.8-A). For the molecular function catalytic activity (GO:0003824) and binding (GO:0005488) were the most prominent categories (Fig 4.8-B).

A total of 36 genes in the organelle category (GO:0043226) and 36 in the cellular (GO:0005623) components (Fig.4.8-C) were the most representative level 2 GO terms in the down-regulated genes data set. A total of 13 pathways were retrieved using Blast2GO version 2.7.0. The most representative pathways included retinol metabolism, steroid hormone biosynthesis, tryptophan metabolism, linoleic acid metabolism, arachidonic acid metabolism, drug metabolism - cytochrome P450, aminobenzoate degradation, caffeine metabolism, fatty acid degradation, carbon fixation in photosynthetic organisms, glyoxylate and dicarboxylate metabolism and oxidative phosphorylation (Table 4.4 – Appendix A).

For the down-regulated genes, 118 transcript sequences were searched against the NCBI nr database using the Blast2GO software with the same e-value cut-off as for the up-regulated genes. About 97 of these, which represent 82.2% of the down-regulated set, obtained BLAST hits.

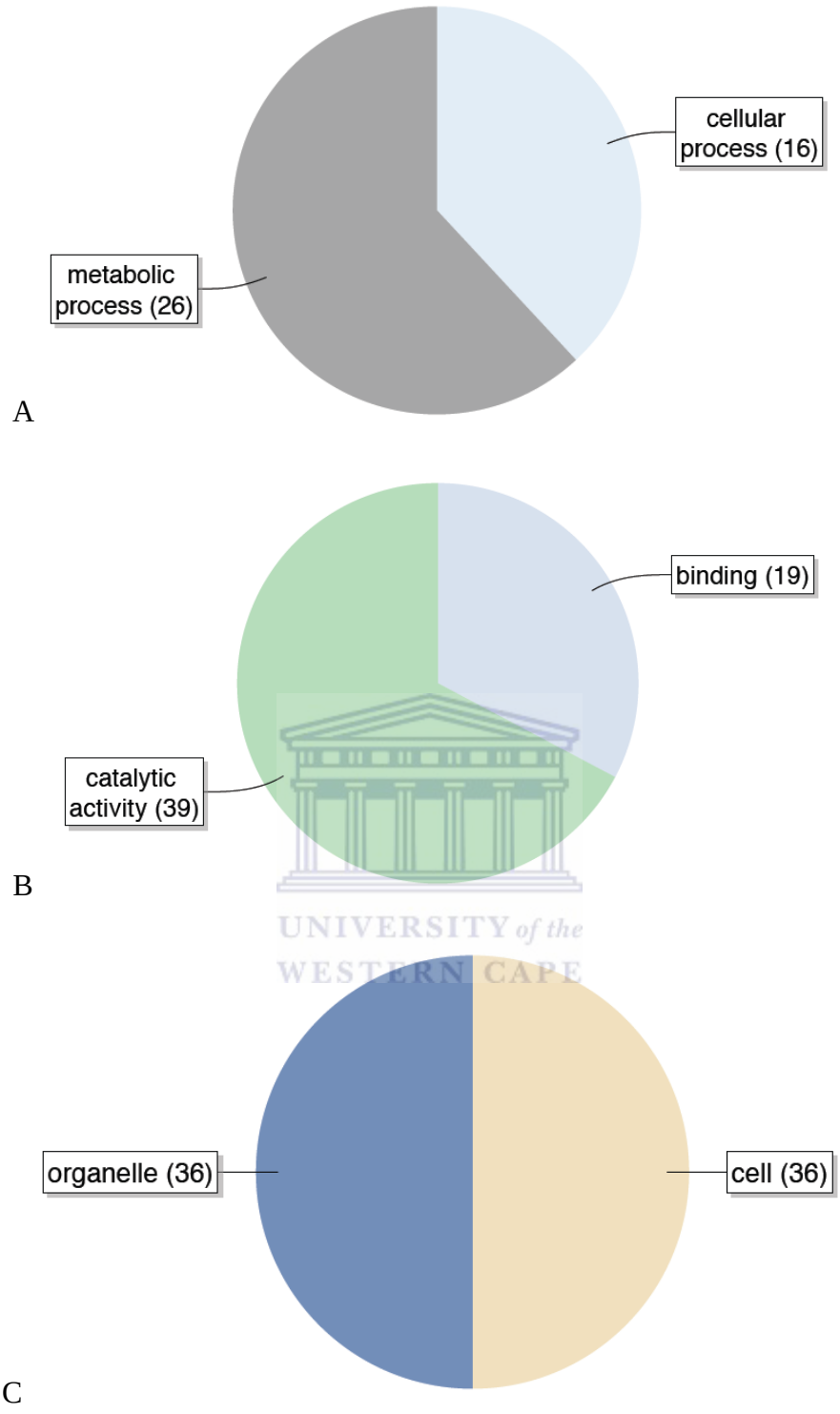


Figure 4.8 GO terms associated for the down-regulated transcripts. A - Biological processes, B - molecular function and C - cellular components level 2 upon infection with *Venturia inaequalis*.

4.6.1 Response of apple after inoculation with *Venturia inaequalis*

The annotation of the differentially expressed genes revealed that many genes involved in plant secondary metabolism are induced such as cytochrome P450 genes as well as transporters among which are ATP-binding cassette (ABC) transporters. Consistent with the spreading invasion of the fungus, many defence genes were activated such as chitinases, glucanases, peroxidases and pathogen/pathogenesis related proteins. Therefore, to understand how many genes were potentially a part of the defence system, the focus was on those involved in signal perception, transcription, stress/detoxification, defence related proteins, transport and secondary metabolites (Fig. 4.9, Table 4.5 – Appendix A). Each of these aspects was discussed in detail from section 4.6.2 onwards.

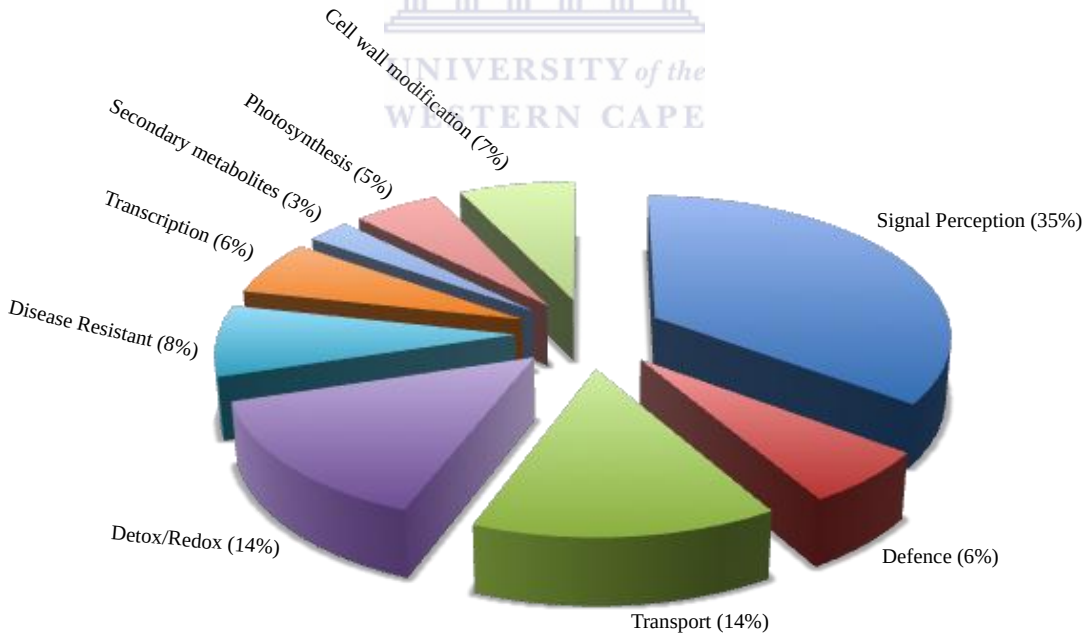


Figure 4.9 Functional categories affected during the fungal infection.

4.6.2 Signal perception and transduction during *Venturia inaequalis* infection

In this study several differentially regulated transcripts encoding for proteins in the signal perception category (35%; Fig. 4.9) were up-regulated. These include receptor-like kinases (RLKs), including a wall-associated receptor kinase-like (WAKL) similar to *Arabidopsis* WAKL2, probable receptor-like protein kinase *at*1g67000, wall-associated receptor kinase-like 10, receptor-like protein 12 and WAKL 2. For example, a transcript encoding a WAK gene and other members of this family were also induced, such as WAKL10. Furthermore transcripts encoding kinesin-related protein 8-like isoform x2, a plant-type cell wall modification protein, was up-regulated. Several leucine rich repeat RLKs (LRR-RLKs) were up-regulated after *Venturia inaequalis* inoculation.

Differentially regulated genes encoding LRR receptors such as leucine-rich repeat receptor-like serine threonine tyrosine-protein kinase sobir1, leucine-rich repeat receptor-like protein kinase *at*1g35710, brassinosteroid insensitive 1-associated receptor kinase 1 and leucine-rich repeat transmembrane protein kinase were identified in this study. Furthermore, leucine-rich repeat transmembrane protein kinase, a close homolog to the FLS2 gene has been up-regulated upon *Venturia inaequalis* inoculation. A number of genes encoding cysteine rich kinases (CRKs) were up-regulated.

4.6.3 Transcription factors involved in signaling

In this study, several different genes corresponding to WRKY transcription factors such as WRKY70, 51, 33, 6 were up-regulated in *Venturia inaequalis*-infected leaves.

Other transcription factor genes found in this study to be induced, were zinc finger and myb 12 transcription factor. These transcripts constitute 6% of the transcription category to be up-regulated.

4.6.4 Detoxification

Several transcripts encoding glutathione S-transferases (GST) and other isoforms of this protein have been found to be up-regulated, an indication that the plant is reacting to oxidative stress. In addition, transcripts encoding a putative elicitor-inducible cytochrome P450 and other members of the same family such as cytochrome P450 82*a*3-like, cytochrome P450 71*a*1-like, cytochrome P450 71*a*4-like, cytochrome P450 71*a*25-like, cytochrome P450 734*a*1-like, cytochrome P450 90*a*1-like, cytochrome P450 716*b*1-like, cytochrome P450 82*a*3-like, cytochrome P450 716*b*1-like were induced. In this study 3 genes expressing cytochrome P450 like-tbp were found to be down-regulated. These transcripts constitute 14% to be up-regulated in this category (Fig. 4.9).

4.6.5 Deployment of defence proteins

A considerable number of transcripts (6%) were differentially up-regulated following apple scab infection. These were PR-1, PR-2 (β-glucanase), class iv chitinase (formerly PR-4), PR-5 (thaumatin-like protein) and PR-9 (peroxidase).

4.6.6 Putative resistance proteins

Several putative R genes (6%) in this study were up-regulated and identified (Fig. 4.9). These include probable disease resistance protein *at*5*g*66900-like, a protein

which belongs to the disease resistance NB-LRR family and is implicated in signal perception of pathogens; disease resistance family protein lrr family, disease resistance response protein 206-like, enhanced disease susceptibility 1 (EDS1), disease resistance protein rga3-like and disease resistance protein rpm1-like.

4.6.7 Transporters during *Venturia inaequalis* infection

Genes up-regulated in this study constitute 14% (Fig.4.9) in this category. These include genes encoding putative pleiotropic drug resistance (PDR)-type ABC transporters, ABC transporter d family member chloroplastic-like, abc transporter i family member 20-like, phosphate transporter, showed significant increases in transcript levels following *Venturia inaequalis* infection. Genes encoding amino acid transporters such as lysine histidine transporter-like 8 were up-regulated. Other genes encoding for transporters such as auxin-, phosphate-, glucose-, sorbitol-, oligopeptide-, peptide nitrate and white-brown complex abc were amongst the induced genes.

4.6.8 Structural defence of the plant such as cell wall modification

The genes up-regulated in this category constitute 7% of this category (Fig. 4.9). Two genes encoding enzymes of phenylpropanoid metabolism were differentially up-regulated upon the infection with the fungus. Up-regulated genes involved in lignin biosynthesis include phenylalanine ammonia lyase (PAL) and lacto-peroxidase.

A gene similar to the invertase/pectin methylesterase inhibitor family was induced and has been implicated in cell wall modification. A number of beta-glucosidase genes have been found to be up-regulated.

Peroxidase, which has been associated with cell wall modification, in particular cell wall lignification, was up-regulated in this study. Another differentially expressed gene similar to a gene encoding a cell wall hydrolase and a cell wall associated partial were down-regulated.

A gene encoding blight-associated protein is induced in the interaction between *Venturia inaequalis* and its host.

4.6.9 Secondary metabolites such as flavonoids to aid in defence

The following genes involved in flavonoid synthesis namely, phenylalanine ammonia lyase (PAL), chalcone synthase, chalcone isomerase, flavonoid 3'-hydroxylase, flavanone 3-hydroxylase and flavonoids 3', 5'-hydroxylase were up-regulated. These constitute 3% (Fig. 4.9) of the total genes up-regulated.

4.6.10 Differential expression of photosynthesis genes

A number of genes putatively involved in chlorophyll A/B binding protein or to the photosystem I and II were all up-regulated. These constitute 5% (Fig. 4.9) of the genes up-regulated in this category. However, one gene expressing ribulose-bisphosphate carboxylase oxygenase small subunit was down-regulated and one gene expressing ribulose-bisphosphate carboxylase oxygenase large subunit was up-regulated.

4.7 Discussion

In this chapter, RNA sequencing technology was used to analyse the transcriptome of young apple leaves of a susceptible Golden Delicious apple cultivar after *Venturia*

inaequalis inoculation and after mock-inoculation at five different time points (0, 2, 4, 8 and 12 Day post inoculation, dpi). Some of the samples' reads mapped with a high percentage to the reference genome and some with a low percentage as can be seen in Table 4.1. In total, an average of 90.2% of the control and 60.9% of infected reads mapped to the apple genome. The reason for the low mapping percentage could be that the 60.9%, which mapped, were apple specific. The remainder of the infected short reads data set could be *Venturia inaequalis* specific and/or contaminants.

The reason for this experiment was to identify responsive proteins to the pathogen and also to observe the progression of the disease over a number of days. Since no biological replicates were included in the experiment, there was no way to gauge how variable each data point is. This was also reflected in the Venn diagram and the MDS-plot. To accommodate this data set, a statistical package namely edgeR (Robinson *et al*., 2013) and limma/voom (Smyth, 2004) were used to estimate transcript abundance. With this package, a comparative analysis between different time points would not be reliable due to the absence of biological replicates. However with this package, the effect of the treatment namely differentially expressed genes were retrieved. The analysis identified 398 differentially expressed genes of which most were up-regulated.

The annotation of these genes involved mapping the gene loci to the corresponding transcript identities. The resulting transcript identities did not have a corresponding coding domain sequence for all of it and the subsequent genomic sequences were retrieved instead for the relevant gene loci from the apple genome. The up-regulated

genes identified were involved in signal perception, transcription, detoxification, defence, resistance, transporters, structural modifications and secondary metabolites.

4.7.1 Signal perception

Plants respond differently to pathogen infection. As a basal defence, plant cells express receptors with a broad range specificity to detect compounds of the pathogen called pathogen- or microbial-associated molecular patterns (PAMPs/MAMPs), leading to a PAMP/MAMP-triggered immunity. This is made possible by pattern recognition receptors (PRRs) such as proteins in the receptor-like kinase family (RLKs), which represent one of the major receptor systems for intercellular signalling pathways in plants. RLKs "sense" external signals at the plasma membrane and initiate signalling cascades via their cytoplasmic protein kinase domains (Nürberger and Kernmerling, 2006). Several RLKs have previously been described to be involved in plant-microbe interactions (Zipfel *et al.*, 2004, 2006). In this study several transcripts encoding RKL's have been up-regulated in order to initiate signalling cascades.

The oligogalacturonides (OGs) receptor WAK (wall associated kinase) is an RLK that has been characterised in relation to the immune responses against *Botrytis cinerea*. Support for this boost in immune response has been showed in the overexpression of WAK1 in *A. thaliana* enhanced resistance to *B. cinerea* (Brutus *et al.*, 2010). He *et al.* (1998) confirmed that induced expression of WAK1 is necessary for *A. thaliana* survival during general pathogen induced plant responses induced by phytohormones (He *et al.*, 1998). Kohorn *et al.* also observed that a dominant allele of WAK2, WAK3cTAP caused ectopic lesions, ROS accumulation, curling leaves, and stunned

growth, which are distinct features of a pathogen infection (Kohorn *et al.*, 2009). The up-regulation of these RLKs in this study might also have similar symptoms mentioned by Kohorn *et al.* (2009) on *Venturia inaequalis* infected apple leaves, indicative of a pathogen infection.

With several induced transcripts encoding resistance (R) proteins, transcripts of this class were predominant in the signal perception mechanisms category. Some of these are members of the nucleotide binding site leucine-rich repeat (NBS-LRR) and cysteine-rich receptor-like protein kinase resistance genes. Leucine-rich repeat RLKs (LRR-RLKs) are among the RLKs used to detect microbial invasion. In this study, the RNA-Seq data detected several LRR-RLKs after *Venturia inaequalis* inoculation. Several LRR receptors were found to be up-regulated as discussed in 4.6.2. The dual specificity kinase present in LRR-RLKs acts on both serine/threonine- and tyrosine-containing substrates. This promotes the activation of plant defence and cell death. Leucine-rich repeat receptors constitute the pattern-recognition receptor (PPR) that determines the specific perception of flagellin (flg22), a potent elicitor of the defence response to pathogen-associated molecular patterns (PAMPs) (Zipfel *et al*., 2004). FLAGELLIN SENSING 2 (FLS2), a transmembrane receptor kinase for bacterial flagellin, is one of the best-characterised PAMP-triggered immunity genes in *A. thaliana*. A close homolog to this gene is up-regulated in resistant soybean inoculated with *Xanthomonas axonopodis* pv. *glycines* and also in peach challenged with *Xanthomonas arboricola* pv. *pruni* (Socquet-Juglard *et al*., 2013). Flagellin-binding to the receptor is the first step in initiating the innate immune mitogen-activated protein (MAP) kinase signaling cascade (MEKK1, MKK4/MKK5 and

MPK3/MPK6), resulting in enhanced resistance against pathogens. The up-regulation of these genes enhances resistance to *Venturia inaequalis* infection.

Cysteine-rich receptor-like kinases (CRKs) are one of the largest RLK groups and play important roles in the regulation of pathogen defence and programmed cell death in *A. thaliana* (Czernic *et al.*, 1999; Chen *et al.*, 2004). CRKs are induced by oxidative stress, pathogen attack such as *Pseudomonas syringae* (Acharya *et al.*, 2007) and *Rhizoctonia cerealis* (Yang *et al.*, 2013) and application of salicylic acid (SA; Czernic *et al.*, 1999; Chen *et al.*, 2003; Chen *et al.*, 2004). A number of genes encoding CRKs were up-regulated in this study. The up-regulation of these genes suggests a putative role for PRR-RLKs in mediating events in the response of plants to *Venturia inaequalis* and their importance for fast pathogen recognition before entrance into plant cells and structures. This suggests that the induction of the genes coding for these proteins is an attempt of the plant to recognise the intruder and to establish a proper defence response.

4.7.2 Transcription

After perception of pathogen associated molecular patterns (PAMPs), genes involved in signalling as discussed in 4.7.1 were activated. This was followed by the activation of several transcriptional regulators such as WRKY members. They play a major role in the regulation of defence responses to pathogens (Rushton *et al*., 1996; Eulgem, 2006). Several different genes corresponding to WRKY transcription factors were up-regulated in this study. These include homologs of *Arabidopsis* WRKY51, 33, 6 and 70. The transcription factors interact specifically with the W box (5'-TTGACC/T-3') (Rushton *et al*., 2010), a frequently occurring elicitor-responsive cis-acting element.

Six different orthologs of *Arabidopsis* WRKY transcription factors were shown differentially expressed following *Venturia inaequalis* infection and the majority of these, such as WRKY70, are clearly associated with defence response (Li *et al*., 2004; Knoth *et al*., 2007). Li *et al.* reported that the overexpression of WRKY70 caused constitutive signaling of salicylic acid. The up-regulation of these transcription factors in this study could function in systemic acquired resistance (SAR).

4.7.3 Detoxification of the host following infection

Oxidative burst is characterised by the rapid generation of reactive oxygen species (ROS) (Lamb and Dixon, 1997), which is indicative of the early events of resistant and susceptible responses to pathogens. ROS species such as O2−, OH•, and H_2O_2 are commonly produced under stress conditions. They are strong oxidizing species that can rapidly attack all types of biomolecules and cause serious damage. It plays an important role in host cell death, which is indicative of a successful infection. Hancock *et al.* (2002) established that ROS act as signalling molecules during host-pathogen interactions (Hancock *et al*., 2002). ROS have dual functions, therefore a controlled generation is of tremendous importance and plants are equipped with several scavenging and anti-oxidant mechanisms for oxidative stress protection. ROS are eliminated directly by several anti-oxidant enzymes such as catalase, peroxidase, and superoxide dismutase, and non-enzymatic antioxidants such as ascorbate peroxidase and glutathione-S-transferase (Pnueli *et al*., 2003). The ascorbate-glutathione cycle is an important anti-oxidant mechanism in plants that consists of a series of reactions resulting in the removal of H_2O_2 (Kuzniak, 2010). These enzymes play a crucial role in the protection of plant cells from oxidative damage at the sites of enhanced ROS generation (Kuniak and Sklodowska, 2001).

Several transcripts encoding glutathione S-transferases (GST) and other isoforms of this protein have been up-regulated. Quite a number of the cytochrome P450 family was up-regulated in this study following the infection with *Venturia inaequalis*. Cytochrome P450's could play roles as antioxidants, UV protectants, detoxification of pollutants, biosynthesis of hormones, but are also involved in basal plant defence against a variety of pathogens (Werck-Reichhart and Feyereisen, 2000). The up-regulation of such a variety of detoxifying genes is indicative of the host reacting to oxidative stress in order to protect the cells against the damaging effects of ROS.

4.7.4 Defence proteins

The synthesis of pathogenesis related proteins are seen as chemical barriers being erected to curb the invading pathogen. This phenomenon is related to age-related resistance to some pathogens. Gusberti *et al.* (2013) found a number of these PR-proteins (β-glucanase, chitinase, endochitinase, thaumatine-like, defensin, oxalate oxidase, and protease inhibitor) encoding genes were found to be down-regulated in old apple leaves (Gusberti *et al*., 2013). In this study a considerable number of transcripts were differentially up-regulated following *Venturia inaequalis* infection. These were PR-1, PR-2 (β-glucanase), class iv chitinase (formerly PR-4), PR-5 (thaumatin-like protein), and PR-9 (peroxidase). Thaumatin-like proteins have homology to permatins, which permeabilise fungal membranes (Van Loon and Van Strien, 1999), which could function in degrading the fungal cell wall. Peroxidases function in strengthening plant cell wall by catalysing lignin deposition in reaction to microbial attack (Van Loon and Van Strien, 1999) as discussed in 4.7.7. Through the strengthening of the cell wall, growth of the fungus is limited due to starvation as a result of reduced nutrient availability (Gusberti *et al*., 2013). Peroxidases have also

been implicated in the regulation of ROS in plants (Kawano, 2003). Furthermore, many of these PR genes are normally involved in the salicylic acid (SA)-dependent pathways (Glazebrook, 2005), such as chitinase and PR-1, were induced following *Venturia inaequalis* infection to play a role in systemic acquired resistance.

4.7.5 Resistance proteins

Effector proteins are secreted by pathogens, which interfere with the host cell function to induce disease. Plant resistance (R) genes recognise these effectors, a gene-for-gene resistance is displayed, which leads to extracellular oxidative burst, induction of salicylic acid signalling pathway, calcium and hydrogen influxes in the cell, and cell death or hypersensitive response (Bolwell *et al*., 2002). Several putative R genes were up-regulated in this study and identified. These include probable disease resistance protein *at*5*g*66900-like, a protein which belongs to the disease resistance NB-LRR family and is implicated in signal perception of pathogens; disease resistance family protein lrr family, disease resistance response protein 206-like, enhanced disease susceptibility 1 (EDS1), disease resistance protein rga3-like and disease resistance protein rpm1-like.

Disease resistance response protein 206-like was observed for the first time in *Pisum sativum* cv. Alaska, a strong resistance response in reaction to the fungus *Fusarium solani*, the bacterium *Pseudomonas syringe* pv. *pisi* and the fungal elicitor chitosan (Fristensky *et al*., 1988). This gene increased resistance to the pea pathogen *Fusarium solani* f. sp. *phaseoli* (Culley *et al*., 1995), which is coupled to the synthesis of lignan and lignin biosynthesis (Davin and Lewis, 2000). This protein expressed in transgenic plants, showed resistance to a broad range of fungi including *Leptosphaeria maculans*

isolates PG2, PG3, PG4, *Rhizoctonia solani* and *Sclerotinia sclerotiorum* (Wang and Fristensky, 2001). In this study, this protein is up-regulated and seems to respond to *Venturia inaequalis* infection, thus broadening its range of resistance.

The enhanced disease susceptibility 1 (EDS1) protein was first observed in a mutant of *A. thaliana*, which was susceptible to *P. parasitica* (Parker *et al*., 1996). The silencing of this gene increased disease resistance in *Arabidopsis* (Aarts *et al*., 1998; Parker *et al*., 1996). The EDS1 gene has been found to be necessary for the functionality and signal transduction of other resistance genes in *Arabidopsis* plants (Parker *et al*., 1996). In this study, the EDS1 gene encoding protein was up-regulated. This finding is in contrast to Gusberti *et al.* (2013) in which EDS1 was down-regulated in both inoculated and uninoculated old leaves at 96 hours post inoculation (hpi; Gusberti *et al*., 2013). The up-regulation of this gene could contribute in rendering the plant susceptible to *Venturia inaequalis*.

An *A. thaliana* homolog for disease resistance protein rpm1-like has been found to be up-regulated in this study. This protein contains NB-ARC domains (van der Biezen and Jones, 1998) and confers resistance to *Pseudomonas syringae* in *A. thaliana* and has been up-regulated upon infection (Grant *et al*., 1995). The up-regulation of this protein in this study is indicative of conferring resistance to the pathogen.

4.7.6 Transporters

Plant ABC transporters and other transporters are membrane-localised proton driven efflux pumps that extrude toxic compounds from the cell. Previous studies in soybean (Eichhorn *et al*., 2006) and tobacco (Stukkens *et al*., 2005) have suggested that the

involvement of pleiotropic drug resistance (PDR)-type ABC transporters in plant resistance ensures traffic of defence-related metabolites. The up-regulation of transporters in this study may indicate an increase in possible mobilisation and translocation of metabolites during the interaction with the biotroph.

4.7.7 Structural defence such as cell wall modification

One of the plant's strategies to ward off a fungal attack is to strengthen the cell wall by lignin deposition. The phenylpropanoid metabolism plays an important role in lignin biosynthesis. Precursor genes to lignin biosynthesis have been induced. Lignin biosynthesis genes namely phenylalanine ammonia lyase (PAL) and lacto-peroxidase were up-regulated upon *Venturia inaequalis* infection, suggesting that apple leaves respond to pathogen invasion by lignin deposition to strengthen the cell wall.

The gene encoding for the invertase/pectin methylesterase inhibitor protein family was up-regulated and has been reported to be playing an important role in basal disease resistance. An increased expression of this gene in pepper upon infection with *Xanthomonas campestris* pv. *Vesicatoria* was reported (An *et al*., 2008). The up-regulation of this gene in this study could also play a role in increasing basal disease resistance.

Genes encoding beta-glucosidase were induced and are known to play a role in cell wall lignification (Escamilla-Trevino *et al*., 2006). It also plays a role in the activation of phytohormones (Kristoffersen *et al*., 2000) and the synthesis of chemical defence compounds (Morant *et al*., 2008). The up-regulation of beta-glucosidase is indicative of the cell wall being strengthened against fungal invasion.

Peroxidase was found to be up-regulated upon infection with *Venturia inaequalis* infection. It has been associated with cell wall lignification. Furthermore, Ye *et al.* (1990) reported its association with systemic resistance of tobacco to blue mould and Allison and Schultz (2004) reported it being up-regulated due to wounding in northern red oak. A gene similar to cell wall hydrolase and partial cell wall associated were differentially down-regulated. The identification of these differentially expressed genes that could be involved in cell wall modification is indicative of major changes, which may be performed during infection at an early stage, either by degrading or lignifying cell walls to avoid the spread of disease into the vascular tissues.

Biotrophic pathogens develop haustoria after successful invasion of the host within infected cells that serve as feeding organs or contribute to the establishment or maintenance of host compatibility (Voegele and Mendgen, 2003). However this characteristic is also displayed in *Venturia inaequalis*, with the exception that it does not infect/invade living cells. Several transcripts have been induced in other biotrophic infections that could be associated with the efficient penetration and growth of the virulent pathogen in the plant host. For example, a transcript encoding a PR gene with similarity to blight-associated protein p12 was highly induced in the compatible interaction. This gene also shares similarity with an *Arabidopsis thaliana* expansin (Kayim *et al*., 2004), which is associated with cell wall loosening during longitudinal growth of cells and may allow efficient penetration of the invading pathogen. A gene with similarity to blight-associated protein p12 was induced in this compatible interaction between *Venturia inaequalis* and *Malus* x *domestica*. A

possible loosening of the cell wall could be implicated, which facilitated the penetration and/or invasion of *Venturia inaequalis* into the host.

4.7.8 Secondary metabolites such as flavonoids to aid in defence

As expected, several genes encoding enzymes related to plant defence such as flavonoids was transcriptionally up-regulated during the infection process. Flavonoids are a class of important secondary metabolites including flavanones, flavones, dihydroflavonols, flavonols, and flavan 3-ols (catechins; Shi *et al*., 2011). They are ubiquitous secondary metabolites in plants that have been considered to function in a wide range of biological processes, including both abiotic and biotic stress responses. It also plays a crucial role in plant defence against pathogens. In this annotated infected apple leaf transcriptome dataset, multiple transcripts encoding most of the known enzymes involved in the main flavonoid biosynthesis pathway were identified in the flavonoid biosynthesis pathway. This analyses showed that several key enzymes involved were up-regulated, including flavanone 3- hydroxylase, chalcone-isomerase, chalcone-synthase (CHS), flavanol synthase, flavonoid 3',5'-hydroxylase and flavonoid 3'- hydroxylase (Fig. 4.11). The up-regulation of these enzymes after *Venturia inaequalis* infection is in agreement with literature, which states that flavonoids are involved in plant defence against pathogens.

4.7.9 Photosynthesis

A number of genes putatively involved in chlorophyll A/B binding protein or to the photosystem I and II were all up-regulated. However, a gene expressing ribulose-bisphosphate carboxylase oxygenase small subunit was down-regulated and a gene

expressing ribulose-bisphosphate carboxylase oxygenase large subunit was up-regulated.

These results are in contrast with previous observations in which a global down-regulation of photosynthesis genes was observed (Bilgin *et al*., 2010).

4.7.10 Metallothioneins

The total absence of metallothioneins from the infected or uninfected leave transcriptome were surprising since these proteins were found in other studies namely Degenhardt *et al.* (2005) and Gusberti *et al.* (2013). Degenhardt *et al.* (2005) speculated that low levels of RuBisCO and the high level of metallothioneins in the leaves of Remo cv renders the plant unattractive for certain biotrophic pathogens. This distribution of the transcripts for RuBisCO and metallothioneins in the leaves resembles old leaves. Gusberty *et al.* (2013) reported an up-regulation of metallothionein proteins in old leaves. In this study, metallothioneins are absent and ribulose-bisphosphate carboxylase oxygenase large subunit was up-regulated. The absence of this protein together with increased RuBisCO levels, is an ideal environment for the fungus to thrive in. A major conclusion of this study is that, in addition to the up-regulation of a number of well-characterised PR proteins, the absence of metallothioneins and the up-regulation of EDS1 gene, which contribute to the susceptibility, seemed to be the contributing factors in rendering the Golden Delicious cv susceptible to infection by the fungus *Venturia inaequalis*.

In general, the up-regulation of these proteins is an indication that the plant responded to the infection by preparing the cells for a potential fungal attack. The genes up-regulated cover all aspects of the spectrum namely, signal perception including

programmed cell death, transporters to pump secondary metabolites, defense, resistance proteins, detoxification proteins that serve as antioxidants to minimize oxidative damage to the cells, strengthening of the plant cell walls to prevent the fungus from spreading to healthy tissue. This is in order to minimize damage to the host and restriction of the invading pathogen.

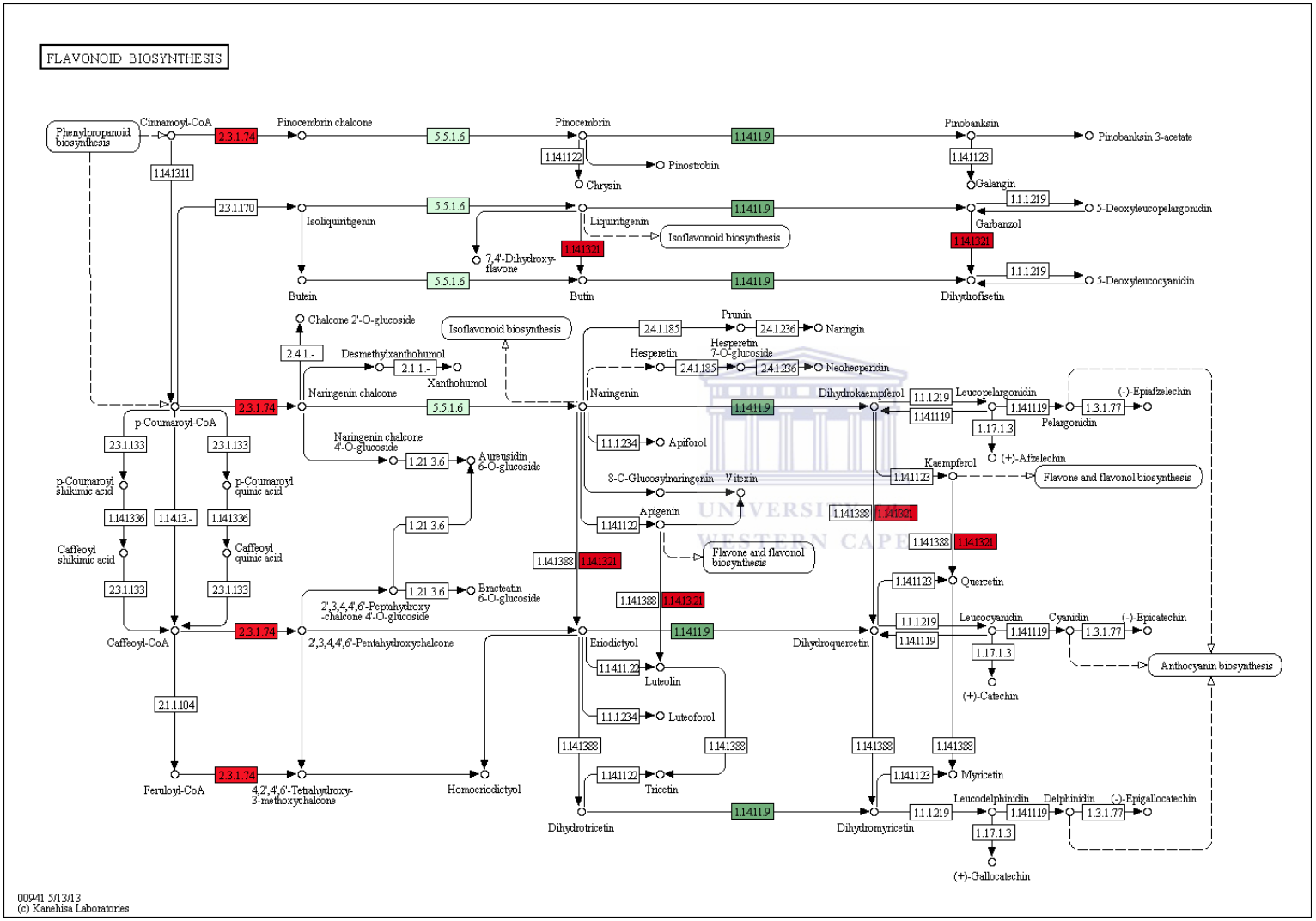


Figure 4.10 Pathway for flavonoid biosynthesis. A number of genes involved in this pathway were up-regulated upon *Venturia inaequalis* infection.

Furthermore, this would ultimately result in a healthier plant and blemish free fruit to the consumer. This results in less pesticide usage and a positive impact on the environment.

In conclusion, the RNA-Seq technique confirmed its suitability for transcriptome studies and for identifying genes differentially expressed in response to a fungal infection. In fact, of the 398 differentially expressed sequences as many as 331 were up-regulated and 67 were down-regulated. Results obtained through this technology could clarify the intricate network of genes involved in the host-pathogen interaction.

Chapter Five - A proteomic analysis of early response in the interaction between *Venturia inaequalis* and susceptible *Malus* x *domestica* (Borkh) cv Golden Delicious

5 Introduction

Apple scab is a seriously damaging disease in apples, caused by the hemibiotroph, *Venturia inaequalis*. The disease renders the fruit being unsightly and unmarketable, which ultimately leads to a loss of revenue. Currently, spraying with fungicides is the main method used by farmers to manage this disease. These spray schedules are costly, in some cases requiring up to 20 applications per year. The extensive use of fungicides is a potential threat to the consumer and this has been raised in environmental awareness campaigns. *Venturia inaequalis* preferentially infects tissues such as young leaves, twigs and fruit. Spores or conidia from *Venturia inaequalis* germinate under favourable conditions on leaf tissue. Upon germination, the germ tube or hypha penetrates the epidermis of the leaf but it does not enter the cell. It stays within the intercellular spaces where it colonises, spreads its hyphae and ruptures the epidermis of the leaf, leaving the leaf exposed to the environment. The fungus behaves as a biotroph by not killing its host. This interaction between the host and the fungus has intrigued many a researcher since the late 1930's (Nusbaum and Keitt, 1938; Nicholson, 1972; Smereka *et al.*, 1987; Koller *et al.*, 1991; Gau *et al.*, 2004; Degenhardt *et al.*, 2005; Schumaker *et al.*, 2008) until now.

Nicholson *et al.* proposed that an esterase-like activity present at the germ tube is responsible for softening the cutin for easy penetration of the cuticle. Koller *et al.* reported that germinating germ tubes and mycelia secrete extracellular cutinases, which play a role in penetration of the cuticle by the pathogen (Koller *et al.*, 1991).

Host-pathogen interaction studies where the response of susceptible and resistant apple cultivars to *Venturia inaequalis* at a molecular level was studied during the last decade. These include transcriptomics (Degenhardt *et al.*, 2005), which is the study of gene expression profiles in an organism (see section 1.7.2) and proteomics approaches (Gau *et al.*, 2004), which is the large-scale study of proteins from an organism, tissue or cell at a given time (see section 1.7.1.1). Gau *et al.* revealed the presence of pathogenesis related proteins (PR), namely PR-2 (ß-1,3-glucanases), PR-4 (chitinase), PR-5 (thaumatin-like protein) and a cysteine-like protease in the apoplast of a susceptible apple cultivar, Elstar. The same study also revealed a constitutive production of the PR proteins in the resistant cultivar, Remo. These findings were in accordance to a study done by Degenhardt *et al.* (2005) who showed that the levels of transcripts encoding a number of proteins related to plant defence (such as β-1,3-glucanase, ribonuclease-like PR10, cysteine protease inhibitor, endochitinase, ferrochelatase, and ADP-ribosylation factor) or detoxification of reactive oxygen species (such as superoxide dismutase) were highly up-regulated relative to the susceptible cultivar, Elstar. They concluded that the constitutive expression of these proteins increased the cultivar, Remo's, resistance to the fungus.

These studies provide valuable information regarding host-pathogen interaction studies between resistant and susceptible cultivars, but the need to understand how the

susceptible cultivar reacts, exists. The measurement of protein expression using proteomics as a tool would provide better indication of cellular activities under biotic stress. This will give more insight into the sequence of events that occurs in the *Malus* x *domestic* cv. Golden Delicious proteome during infection. The recently sequenced apple genome (Velasco *et al*., 2010) has paved the way for downstream protein identification steps by mass spectrometry analysis.

The chapter aimed at identifying pathogen response (PR) proteins in apple leaf extracts using 2-DE and MS. The data obtained would deepen our understanding of how the plant responds to the fungus in order to combat the infection.

5.1 Inoculation of apple seedlings

Young apple seedlings were inoculated with a local *Venturia inaequalis* strain and collected as described in section 2.5.

5.2 Total protein extraction from apple leaves and quantification

Total protein was extracted following a modified protocol as described in section 2.9.1. Extracted leaf proteins were quantified using a modified Bradford assay as described in section 2.10 using BSA as standard.

5.3 1-DE and 2-DE

One dimensional gel electrophoresis of 20 µg of leaf protein extracts was carried out to evaluate both the quality and the loading quantities of the extracts prior to 2-DE

(see section 2.11 for more detail). For 2-DE, mini format SDS-PAGE gels were run for use in comparative gel analysis as described in section 2.12.

5.4 Comparative analysis of 2-DE gels

Comparative analysis of protein expression on 2-DE gels between uninfected and infected treatment groups was carried out using Melanie 7.0 (Genebio, Geneva, Switzerland) as described in section 2.13.1.

5.5 Protein identification by MALDI TOF-TOF MS

Protein spots of interest were reduced, alkylated and trypsin digested as described in section 2.13.2.1. Dried peptides were dissolved and injections were made for nano-LC chromatography connected to a LTQ Orbitrap Velos™ mass spectrometer (Thermo Scientific, Bremen, Germany) as described in section 2.13.2.2.

5.6 Bioinformatics analysis

Bioinformatic analysis of the proteins of interest was performed following the pipeline as described in section 2.13.2.5.

5.7 Sample infection and collection: Post inoculated

Young Golden Delicious apple seedlings infected with the *Venturia inaequalis* strain showed no signs of infection four days post-inoculation (dpi) (Figure 5.1). Leaves for the extraction of total proteins were collected as described in section 2.5.2.

a  b 

F**igure 5.1** Young leaves showing no infection 4 days post-inoculation. Leaf (a) is the control mock inoculated and (b) is infected with an inoculum.

5.8 Protein extraction from 4 dpi young seedling leaves

Golden Delicious apple seedling leaves were used as source of plant material. The plant material was immediately flash frozen in liquid nitrogen to minimize protein degradation and subsequently stored at -20°C. Total protein was extracted from apple leaves using the phenol/SDS precipitation method as described in section 2.9.1. Extracted proteins were quantified using the Bradford assay as described in section 2.10 with leaves yielding approximately 6-8 μg/μL. Following separation on one-dimensional (1-DE) PAGE, clear distinct bands were visible (Figure 5.2) indicating the presence of intact proteins.

5.9 1-DE protein profiles of the fungal infected and uninfected apple leaves

One-dimensional (1-DE) gel electrophoresis as described in section 2.11 was carried out to evaluate the quality and loading quantities of the protein extracts prior to second-dimensional (2-DE) gel electrophoresis. Figure 5.2 shows Coomassie Brilliant Blue (CBB) stained 1-DE profiles of apple leaf proteomes for the infected and control. Lane 1 shows the molecular weight markers. Lanes 2-4 (control) and lanes 5-

7 (infected) show protein profiles from three biological replicate extractions of the leaf tissues. Each lane was loaded with approximately 20 µg of total protein. The quality of the leaf protein extracts were good, showing no visible signs of streaking and protein degradation. The biological replicates, lanes 2-4 (control) and 5-7 (infected) also showed high similarity in terms of protein expression, abundance and banding patterns. This suggests that protein preparation was reproducible between independent extractions. When comparing protein loading across the six biological replicates (Lanes 2-7) for both control and infected extracts, it was observed that protein load was relatively uniform.

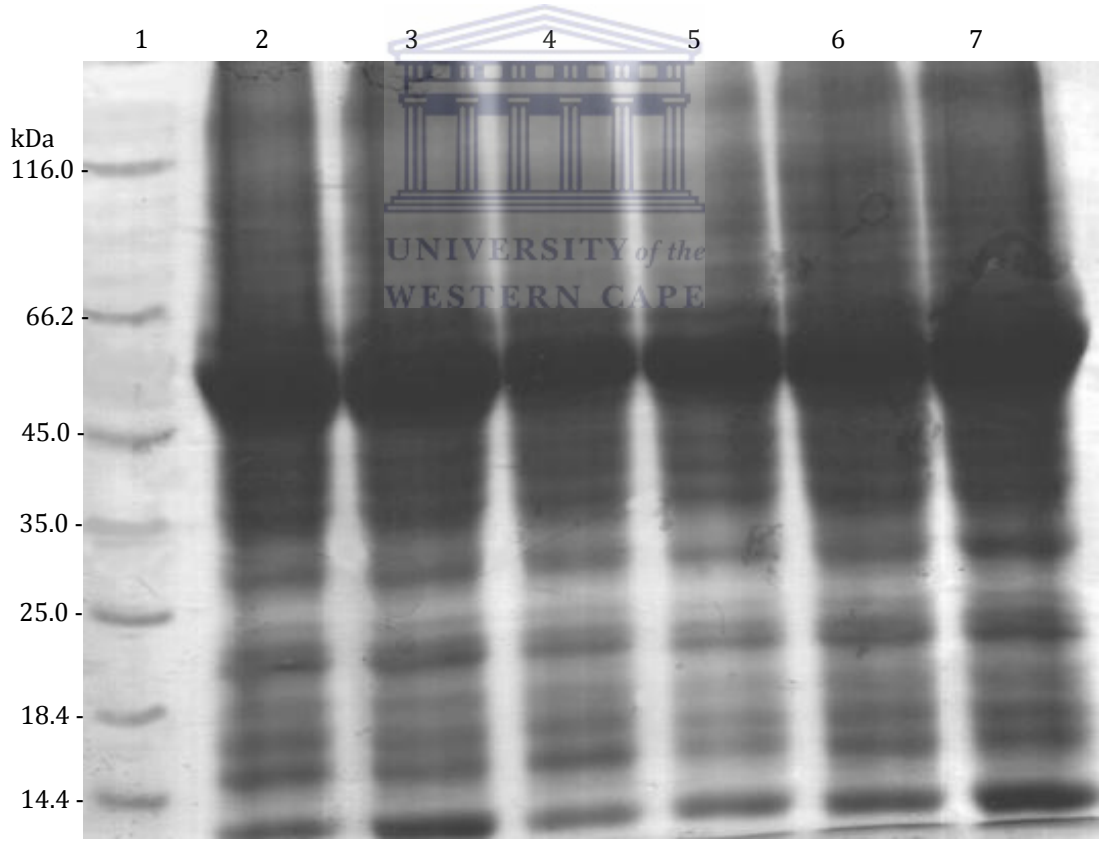


Figure 5.2 1-DE PAGE profile of total protein extracted from Golden Delicious apple leaves. Lane 1 shows the protein ladder, Lanes 2-4 uninfected (control) and lanes 5-7 fungal infection (treatment) leaves. Each lane has 20 µg of total protein.

5.10 Two Dimensional protein profiles of the apple leaf tissues

Three biological replicates of protein extracts from apple leaf tissue (Lanes 2-7, Figure 5.2) were selected for further separation using 2-DE analysis as described in section 2.12, and 250 µg of total protein extracts were loaded on the IPG strips. Mini (7 cm) gels using IPG strips of pH range 4-7 were used. The pH range 4-7 was used because most of the soluble proteins fell within that pH range. It was observed that spot resolution and abundance between three biological replicate gels of each of the leaf extracts were uniform. This indicated that 2-DE was reproducible between different samples within an experiment. Figure 5.3 illustrate representative 2-DE gels of apple leaf proteome of control (A) and infected seedling leaves (B) respectively. In general, it was observed that the control seedlings leaf proteomes shared common spot profiles compared to those coming from treated material. In the 2-DE proteome profiles of the two tissues, the MW range between 14 - 116 kDa.

5.11 Comparative analysis of biotic stress responsive proteins

The main objective of this chapter was to identify stress responsive proteins between the leaf proteomes of uninfected and infected apple leaves. Total soluble proteins obtained by a phenol/SDS precipitation method as described in section 2.9.1 from leaves of Golden Delicious cv. population were first separated on 7 cm long IPG strips across a pH gradient from 4-7 (Figure 5.3). The gels were scanned and imaged using a Molecular Imager PharosFX Plus System (BIO-RAD). Comparative analysis of the stress responsive proteins between the control and infected samples was performed using Melanie 7.0 (Genebio) as described in section 2.13.1. Three biological and three technical replicates of protein extractions were performed, producing a total of 9 independent protein samples for each treatment, all of which

were characterised by 2-DE gel (Figure 5.3). The inclusion of biological replicates in comparative proteomic studies is important for reducing the chances of detecting non-reproducible protein expression differences (artefacts) between experiments. For this reason, all spots included in this analysis were reproducibly expressed amongst the three biological replicates of each treatment group (control and infected). Figures 5.4 and 5.5 showed the Melanie 7.0 (Genebio) analysis gels for control and treatment leaf experiments respectively. In each experiment, three biological and three technical replicate gels are shown per treatment group. For both experiments, the 2-DE protein profiles were reproducible between the gels within each treatments group. The 2-DE gels were of high quality and resolution with no apparent streaking being observed on the gels. The total number of reproducible CBB stained spots per treatment group (control and fungal treatment), are shown in Table 5.1. The gels obtained from the leaf tissues of infected and control seedlings showed hundreds of proteins separated according to their molecular weight (MW) and isoelectric point (*pI*), thus providing an overall picture of the changes that occurred in the plant proteome following the penetration by the fungus (Figure 5.3). The analysis of the proteomes using the Melanie 7.0 (Genebio), followed by visual confirmation, revealed a number of spots showing qualitative and/or quantitative differences between the plant-pathogen interactions (infected) and the control plants.

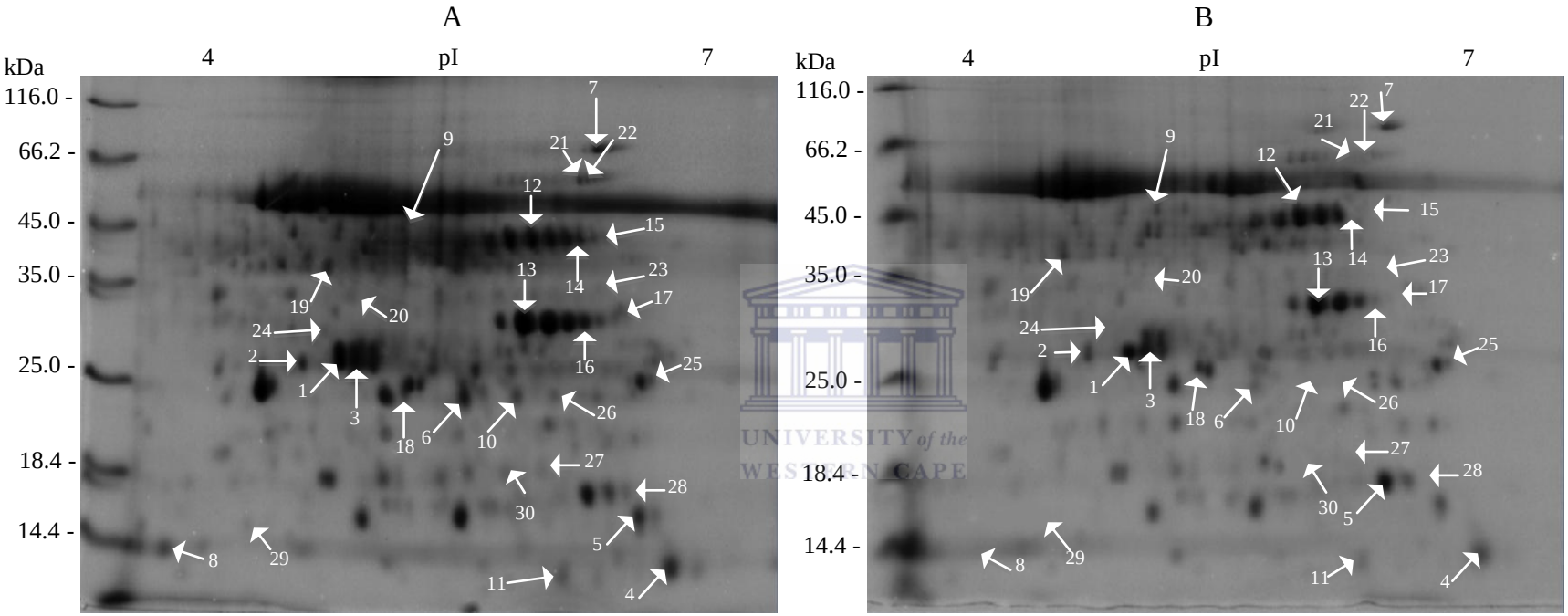


Figure 5.3 Proteome-level investigation of *Malus* x *domestica* cv. Golden Delicious leaves in response to the inoculation of a *Venturia inaequalis* isolate. Representative mini format gels of control (A) and *Venturia inaequalis* infected (B) leaf proteome showing general protein expression within the two treatment groups. Total proteins were extracted at 4 days post inoculation.

Table 5.1 Differentially expressed protein spot analysis using Melanie Software programs

Sample number	*Max*	*Control*	*Infected*	*Anova*	*Fold Expression*	
					Control	*Infected*
1	3,42144	2,19766	3,42144	1,24E-05		1,22378
2	1,42944	0,755327	1,42944	1,86E-06		0,674113
3	3,04648	1,80507	3,04648	0,00546388		1,24141
4	2,78685	1,95545	2,78685	4,86E-04		0,8314
5	2,35874	1,43937	2,35874	1,15E-12		0,91937
6	2,06412	2,06412	0,636858	1,46E-11	1,427262	
7	1,75918	0,764398	1,75918	6,09E-05		0,994782
8	1,29181	1,29181		1,81E-16	1,29181	
9	1,09638	1,09638	0,258611	0,00571502	0,837769	
10	0,852577	0,852577		2,37E-19	0,852577	
11	1,31778	0,7898	1,31778	1,49E-11		0,52798
12	1,95322	1,42606	1,95322	0,00183162		0,52716
13	3,12616	2,58316	3,12616	6,85E-04		0,543
14	0,58295	0,58295		1,00E-10	0,58295	
15	0,426351	0,426351		4,80E-06	0,426351	
16	0,955671	0,955671	0,455156	1,41E-08	0,500515	
17	0,47904	0,47904		3,01E-19	0,47904	
18	0,658268	0,658268		1,26E-09	0,658268	
19	0,811406	0,811406	0,275793	0,0205688	0,535613	
20	0,446911	0,446911		3,11E-06	0,446911	
21	0,222072	0,222072		1,17E-15	0,222072	
22	0,154943	0,154943		3,56E-18	0,154943	
23	0,157986	0,157986		2,02E-04	0,157986	
24	0,347353	0,347353		6,41E-11	0,347353	
25	0,108516	0,108516	0,117613	3,7E-13	0,109466	
26	0,304157	0,304157		6,19E-15	0,304157	
27	0,286666	0,286666		2,70E-06	0,286666	
28	0,37674	0,37674		1,29E-17	0,37674	
29	0,302138	0,302138		6,05E-12	0,302138	
30	0,124039		0,124039	5,86E-08		0,124039

The pattern of differentially expressed protein spots between the control and treatment can be represented in a quantitative and qualitative analysis set created by the Melanie 7.0 software (Genebio). A description of each of the analysis sets is given below.

5.12 Analysis sets for leaf response proteins

5.12.1 Qualitative differential protein expression

In the qualitative analysis, a qualitative spot is defined as a spot that was detected in either the control or treatment groups but not in both. Therefore, qualitative spots can

either be (i) spots that were induced by the fungus and are thus only present in the fungal treatment group; or (ii) spots that completely disappeared after the fungal treatment and are therefore only present in the control treatment group. The *Venturia inaequalis* strain produced the following changes, compared to the un-inoculated control: 1 spot present in treatment but absent in control, 16 spots were absent in treatment but present in the control.

5.12.2 Quantitative differential protein

A spot was considered to be quantitatively differential if, in comparison to the other condition, it was differentially accumulated/reduced by at least ≥ 0.5 fold. It was observed that spots having a spot volume value of less than 0.5 fold difference, were very faint on the gel and therefore not included in the analysis. To exclude any biases such as technical errors, biological variations in plant material regarding differentially expressed spots, three biological and three technical replicate gels per treatment were included in this analysis (Figures 5.4 and 5.5). The fold changes of 2-DE differential spots obtained from the infected compared to the control, are reported in Table 5.1. The total numbers of differential spots that fulfilled these criteria are shown in Table 5.1. For 13 of these, the difference was statistically significant (**ANOVA**, $P < 0.05$).

In Table 5.1, 9 protein spots were up-regulated and 4 protein spots were down-regulated in the Golden Delicious cv. proteome based on the criteria ($P < 0.05$ and fold ≥ 0.5; Table 5.1). This finding indicates that the loss of resistance to the apple scab disease pathogen *Venturia inaequalis* in *Malus* x *domestica* (Borkh) cv. Golden Delicious is associated with substantial modifications in the plant interaction proteome. Some proteins were identified in more than one spot, although they were

excised from the same gel. For example, spots 2 and 3 were identified as triose phosphate isomerase, spots 4 and 24 as ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, spots 6 and 10 as Photosystem II PsbP, spots 9, 14, 15 and 26 as ATPase, AAA-type core, spots 13 and 17 as Photosystem II PsbO, spots 21 and 22 as Chaperonin Cpn60, spots 5 and 28 as Major latex protein domain, respectively. The theoretical and experimental masses and p*I*s were similar for the majority of the identified protein spots. This may be due to the presence of different protein isoforms due to post-translational modifications (PTMs), or multigene families, products of proteolytic activities, the presence of multiple subunits of a single protein and/or the chemical modification of proteins during sample preparation degradation (Yan *et al.*, 2006; Albertin *et al.*, 2009; Gao *et al.*, 2009).

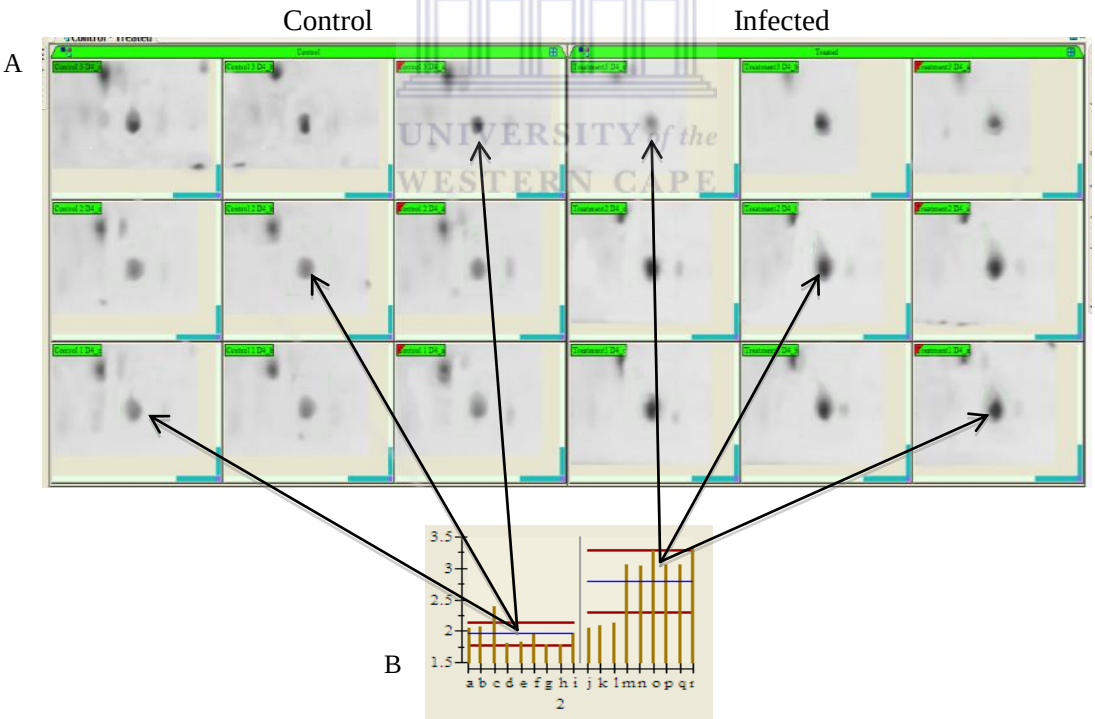


Figure 5.4 Spot 4 shows up-regulation in infected sample. Panel (A) shows control and infected gels with spot being labelled. Panel B shows histogram with up-regulation indicated.

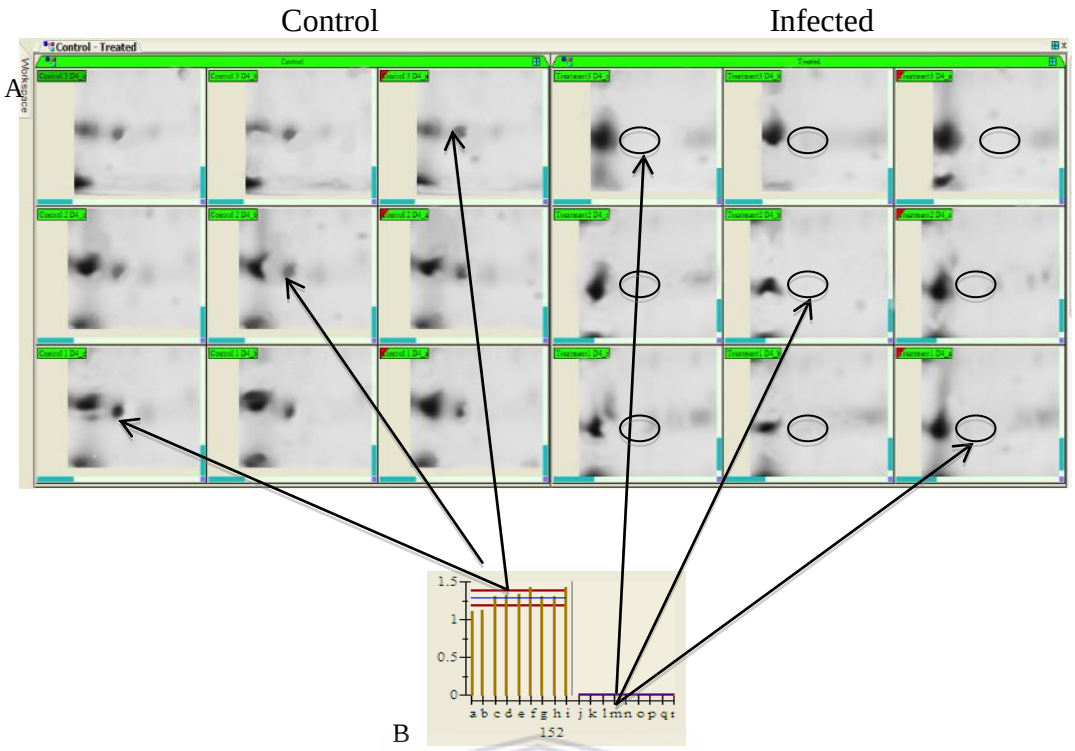


Figure 5.5 Spot 8 showing absent/present in gel. Panel (A) indicates the presence of spot 8 in control but absent in proteins from infected plants. Ovals in gel image depict position where spot is absent. Panel B shows histogram indicating presence or absence of spot.

5.13 Identification of differentially expressed protein spots

Only the spots whose abundance shows expression in the comparison between the infected and control are reported here. Spots showing quantitative or qualitative differences (Figures 5.4 and 5.5) were subjected to Orbitrap-MS and LTQ-MS/MS respectively. Thermo Proteome Discoverer 1.3 (Thermo Scientific, Bremen, Germany) were used to identify proteins via automated database searching (Mascot, Matrix Science, London, UK) for all tandem mass spectra against the Apple peptide database obtained from the IASMA website (http://genomics.research.iasma.it/). This resulted in 28 hits out of the 30 spots analysed, complying with the following parameters of confidence: (i) similar MW, p*I*; (ii) percentage sequence coverage (sequence coverage above 5% is considered good), sequence coverage (above 24% is

significant; Table 5.2). The majority of the identified proteins showed a significant score above 24%, which is considered significant, according to the Mascot search engine. Twenty-seven proteins of 28 matched to homologs in the apple peptide database (section 2.13.2.4). Protein spot 24 found a homolog in the NCBI non-redundant (nr) database. Furthermore, the sequences of these 28 spots were submitted to the NCBI nr database via Blast2GO® version 2.7.0 (Conesa *et al*., 2005) (www.blast2go.org; Table 5.2) and manually annotated using the apple functional annotation database obtained from Genome Database for Rosaceae (GDR) www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0.

Table 5.2 Differentially expressed proteins identified by MS and MS/MS

Spot Nr[a]	*Accession (Malus x domestica genome)*[b]	*Protein name and species*	*NCBI gi| no.*[c]	*Expression change*[d]	*Score*[e]	*Coverage*[f]	*Unique Peptides*[g]	*Experimental Mr/pI*[h]	*Theoretical Mr/pI*[i]	*Location*[j]
Energy										
2	MDP0000694943	Triose phosphate isomerase *Malus* x *domestica*	gi|390098824	Up	3764.29	70.08	1	27.30/4.80	27.31/5.76	Cyto
3	MPD0000152242	Triose phosphate isomerase cytosolic isoform-like protein *Capsicum annuum*	gi|390098824	Up	1113.50	49.21	11	27.30/5.20	27.33/6.02	Cyto
4	MDP0000597996	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit *Malus* x *domestica*	gi|183448180	Up	64.20	4.42	1	14.0/6.80	25.09/6.30	Cyto
6	MDP0000361338	Photosystem II PsbP, oxygen evolving enhancer protein 2 *Malus* x *domestica*	gi|11134156	Down	1380.18	49.33	8	20.0/5.70	32.43/9.44	Cpl
8	MDP0000524809	Ribulose-1,5-bisphosphate carboxylase small subunit *Malus* x *domestica*	gi|403160	*	430.81	53.07	4	14.0/4.10	20.09/9.05	Cpl
9	MDP0000944409	ATPase, AAA-type, core *Malus* x *domestica*	gi|225434859	Down	568.05	36.65	5	45.0/5.20	51.70/6.19	Cpl
10	MDP0000361338	Photosystem II PsbP, oxygen evolving enhancer protein 2 *Malus* x *domestica*	gi|11134156	*	532.77	41.00	6	20.0/5.80	32.43/9.44	Cpl
12	MDP0000223905	ATPase, AAA-type,core *Malus* x *domestica*	gi|225434859	Up	10308.74	60.16	17	45.0/5.90	54.21/6.28	Cpl

13	*MDP0000248920*	*Photosystem II PsbO, manganese-stabilising Malus x domestica*	*gi|224084209*	*Up*	*1153.07*	*57.53*	*3*	*30.0/5.9*	*35.06/6.08*	*Cpl*
14	MDP0000223905	ATPase, AAA-type, core *Malus* x *domestica*	gi|225434859	*	3738.34	51.83	8	40.0/6.2	54.21/6.28	Cpl
15	MDP0000944409	ATPase, AAA-type, core *Malus* x *domestica*	gi|225434859	*	90.95	5.93	1	40.0/6.5	51.70/6.19	Cpl
16	MDP0000245346	Phosphoglycolate phosphatase-like *Glycine max*	gi|356568529	Down	346.58	16.26	5	30.0/6.4	44.60/4.47	Cpl
17	MDP0000858039	Photosystem II PsbO, manganese-stabilising *Malus* x *domestica*	gi|224084209	*	272.75	29.52	1	35.0/5.90	35.03/5.74	Cpl
19	MDP0000193729	NAD-dependent epimerase/dehydratase *Malus* x *domestica*	gi|224082634	Down	544.60	33.75	6	35.0/5.0	43.70/7.16	Cpl
23	MDP0000244771	Sedoheptulose-bisphosphatase *Malus* x *domestica*	gi|350538149	*	490.41	18.48	6	33.0/6.40	83.09/6.35	Ncl
24	379062075	Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit, partial (chloroplast) *Eriobotrya deflexa*	gi|379062075	*	281.73	22.42	4	30.0/4.90	50.42/5.96	Cyto
26	MDP0000223905	ATPase, AAA-type, core *Malus* x *domestica*	gi|225434859	*	834.88	48.17	9	25.0/6.10	54.21/6.28	Cpl
Protein destination and storage										
18	MDP0000150162	Chaperonin 21 *Malus* x *domestica*	gi|255550363	*	391.03	45.88	6	25.0/5.40	29.31/8.68	Cpl
21	MDP0000631455	Chaperonin Cpn60 *Malus* x *domestica*	gi|224104681	*	3218.91	45.90	3	60.0/6.20	61.94/5.15	Cpl

22	*MDP0000631455*	*Chaperonin Cpn60 Malus x domestica*	*gi|224104681*	*	*3527.07*	*46.08*	*3*	*60.0/6.20*	*61.94/5.15*	*Cpl*
Nucleotide metabolism and transport										
29	MDP0000322880	Nucleoside diphosphate kinase *Malus* x *domestica*	gi|19570344	*	524.32	44.31	7	15.0/4.5	18.66/7.95	Cyto
Disease/defence										
5	MDP0000427722	Major latex protein domain *Malus* x *domestica*	gi|224108133	Up	926.09	79.08	7	16.0/6.2	16.90/4.97	Cyto
28	MDP0000427722	Major latex protein domain *Malus* x *domestica*	gi|224108133	*	149.25	43.79	3	16.0/6.50	16.91/4.97	Cyto
Protein synthesis										
27	MDP0000932838	Regulator of ribonuclease activity A *Malus* x *domestica*	gi|255576633	*	161.76	19.88	1	17.0/6.90	17.54/5.15	Cyto
30	MDP0000295888	Translation elongation factor IF5A *Malus* x *domestica*	gi|357973570	**	331.61	44.38	5	17.0/5.70	17.54/5.46	Cyto
Redox/Response to stress										
1	MDP0000253074	Abscisic acid stress ripening protein *Malus* x *domestica*	gi|292494382	Up	718.05	50.21	5	25.0/5.0	25.66/6.01	Ncl
7	MDP0000687238	Chloroplast HSP70 *Malus* x *domestica*	gi|124245039	Up	2852.04	43.96	3	62.0/6.20	78.53/5.01	Cpl

11	*MDP000068641 9*	*Thioredoxin Malus x domestica*	*gi|225457807*	*Up*	*375.37*	*30.56*	*1*	*10.0/6.0*	*19.63/8.90*	*Cpl*
Other										
20	No Hit									
25	No Hit									

Proteins were grouped into functional categories according to Bevan *et al*. (1998)
a Spot numbers as illustrated in Figure 5.3
b Apple genome accession numbers
c NCBI accession numbers
d Expression changes of protein spots as measured by Melanie software. Up – increase in spot volume following fungal inoculation, down – decrease in spot volume following fungal inoculation. * Indicates that spots under review were present in the control group. **Indicates that spot under review was present only in the treatment group. All reported changes were statistically significant using One-way ANOVA ($p<0.05$).
e Score - Mascot score, anything above 24 is considered significant
f Coverage - The % sequence coverage of the protein as detected by the MS. Sequence coverage of 5% is considered good.
g Unique peptides - Total nr of peptides detected for the protein that is unique to the protein.
h Experimental Mr (kDa) and p*I* of the proteins we estimated from the 2D gel images (Figure 5.4).
i Theoretical Mr (kDa) and p*I* as calculated using Compute pI/Mw tool available on ExPASy (http://web.expasy.org/compute_pi).
j Subcellular location of the proteins as predicted by TargetP version 1.1 http://www.cbs.dtu.dk/services/TargetP/ and http://wolfpsort.org/
Abbreviations: cyto – cytoplasm, clp – chloroplast, ncl – nucleus

5.14 Functional classification of differentially expressed proteins

Expressed proteins identified following mass spectrometry analysis were assigned to functional categories as described in Bevan *et al.* (1998), using a combination of similarity searches on UniProt and literature resources. This knowledge would assist in building knowledge of cellular processes that these proteins might be involved in. This would elucidate the metabolic pathways involved in the host-pathogen interaction. The putative functions of the identified apple leaf proteins (Table 5.2) were assessed by a combination of similarity searches on the National Centre For Biotechnology (NCBI) non-redundant (nr) database via Blast2GO® version 2.7.0 (Conesa *et al*., 2005) and Genome Database for Rosaceae (GDR) functional annotation database of the *Malus* x *domestica* apple genome www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0, Universal Protein Sequence database (www.uniprot.org/) and literature sources.

All the 28 positively identified protein spots (Table 5.2) were successfully classified into six broad functional categories following Bevan criteria (Bevan *et al.*, 1998). The functional categories associated with identified proteins are (i) carbohydrate metabolism/energy, (ii) protein destination and storage, (iii) nucleotide metabolism and transport, (iv) disease and defence, (v) protein synthesis and (vi) redox/response to stress. The functional categories and proteins in each respective class are listed in Table 5.2 while a graphical representation of the distribution of proteins in each class is illustrated in Figure 5.6.

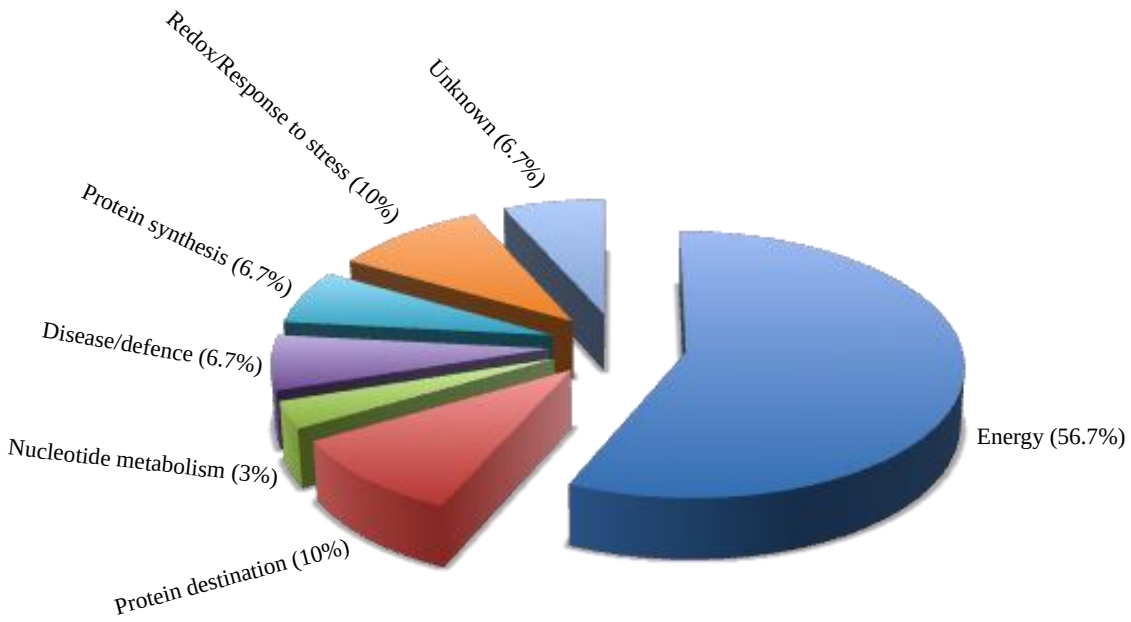


Figure 5.6 Functional classification of the MALDI-TOF and MALDI-TOF-TOF MS identified proteins from infected apple leaves. Numbers indicated in brackets represent the proportion of proteins within each functional category expressed as a percentage of the 30 MALDI-TOF MS positively identified protein spots.

5.15 Functional categories of host proteins

As presented in Table 5.1, the proteins identified in this study as being significantly ($P < 0.05$) affected by *Venturia inaequalis* challenge were grouped according to their known intracellular functions (Figure 5.6). A large proportion (56.7%) of the proteins are functionally associated to photosynthesis and carbon metabolism and fall into the energy category. The second largest group of proteins (10%) was identified as those involved in (i) protein destination and storage and (ii) redox/response to stress. Other categories are comprised of proteins involved in protein synthesis and disease/defence (6.7%)

respectively, and the final group belongs to the nucleotide metabolism and transport (3%). Seven percent of the responsive proteins are unknown.

5.15.1 Carbohydrate metabolism

A number of proteins identified in this study (56.7%) are known to be energy related (Figure 5.6). These include photosynthesis associated proteins such as triose phosphate isomerase (TIM; spots 2 and 3), ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (RuBisCO; spot 4 and 24), ribulose-1,5-bisphosphate carboxylase small unit (spot 8), sedoheptulose-bisphosphatase (spot 23), photosystem II PsbP oxygen-evolving enhancer protein 2 (spots 6 and 10), photosystem II PsbO manganese-stabilising (spots 13 and 17), NAD-dependent epimerase dehydratase (spot 19), ATPase AAA type core (spots 14, 15 and 26) and phosphoglycolate phosphatase-like (spot16; Figure 5.3 and Table 5.2).

5.15.2 Protein destination and storage

Three proteins that fall in the protein destination/storage category were identified. These constitute 10% of the proteins identified in this study. They are Chaperonin 21 (spot 18) and Chaperonin Cpn60 (spots 21 and 22). In this study, one member of the sHSPs *viz.* HSP21/Cpn21 (spot 18) was identified and was found to be absent or degraded in the stressed proteome. The other member of the chaperonins family of HSPs *viz.* HSP60/Cpn60 (spots 21 and 22) is also absent or degraded in the proteome of the infected leaves.

5.15.3 Redox/response to stress

Three proteins belonging to this category were identified. They make up 10% of all the proteins identified in this study. These include abscisic acid (ABA) stress and ripening (ASR) (spot 1), chloroplast Hsp70 (spot 7) and thioredoxin (spot 11). Two of these proteins namely ASR and Hsp70 are stress response proteins.

5.15.4 Nucleotide metabolism and transport

One protein spot 29, nucleoside diphosphate kinase (NDK) was identified. This NDK is known to be involved in the nucleotide metabolism category and constitute 3% of all the proteins identified. This protein was found to be absent or degraded in the proteome of the infected apple leaves.

5.15.5 Protein synthesis

Two proteins (spots 27 and 30; Figure 5.3; Table 5.5) represented the protein synthesis group thus constituting 6.7% of all the identified proteins. The proteins identified were the regulator of ribonuclease activity (spot 27) and the translational elongation factor IF5A (spot 30). In this study RraA protein is absent from the infected apple leaf proteome whereas the IF5A protein is present.

5.15.6 Disease/defence

Two protein spots in this category were identified, which constitutes 6.7% of the proteins analysed. These proteins were identified as major latex protein (MLP) domain (spots 5

and 28). In this study we found two isoforms of MLP, one showed an increase in abundance and the other is absent/or silenced in the proteome of the infected apple leaves.

5.16 Subcellular localisation of identified proteins

The subcellular localisation of a protein provides important information about its physiological function (Fang *et al.*, 2012). Subcellular localisations of the identified apple leaf proteins were predicted using a combination of TargetP version 1.1 (http://www.cbs.dtu.dk/services/TargetP/; (Emanuelsson *et al.*, 2007), WoLF PSORT http://wolfpsort.org/ (Horton *et al.*, 2007) and literature sources. The subcellular localisation of each of the positively identified proteins is given in Table 5.5 while a graphical representation of the total number of proteins in each subcellular location is shown in Figure 5.7. Most of the identified proteins are located in chloroplasts (56%), the cytoplasm (30%), nucleus (7%) and unknown (7%; Figure 5.7).

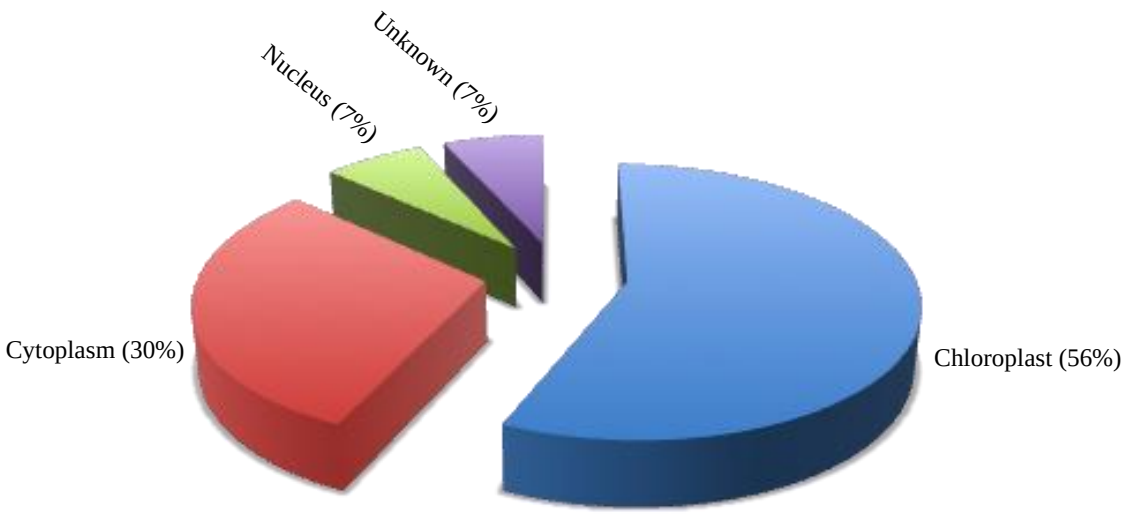


Figure 5.7 Subcellular localisation results of identified apple leaf proteins. Subcellular localisations of the identified apple leaf proteins were predicted using a combination of TargetP version 1.1 (http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson *et al.*, 2007), and http://wolfpsort.org/; Horton *et al.*, 2007) and literature sources. Numbers indicated in brackets represent the proportion of proteins within each subcellular compartment expressed as a percentage of the 30 spots that were positively identified by a combination of MALDI-TOF and MALDI-TOF-TOF MS and database searches.

5.17 Discussion

5.17.1 Protein extraction

Young apple leaf tissue yielded approximately 6-8 µg/µL protein. The low protein concentration in addition with contamination by pigments, lipids and phenolics pose a major challenge for 2-DE analysis of apple leaf protein (Granier, 1988). Low quantities of protein were also found in olive and grape leaves (Wang *et al.*, 2003; Basha *et al.*, 2009). Chaotropes, detergents and ampholyte combinations were used to enhance the solubility and resolution of apple leaf proteins, and to minimize interference from leaf

metabolites. The solubilisation efficiency depends not only on the nature of leaf proteins, but also on the leaf morphology, lipid and pigment content, and the sample preparation prior to final solubilisation (Wang *et al.*, 2003).

Total protein extraction involves mechanical disruption of the plant tissue under liquid nitrogen to minimize proteolysis and other modes of protein degradation. The use of TCA and acetone in combination, aids in precipitating proteins and removal of contaminants during preparation for 2-DE. The advantage of this protocol is that it reduces the exposure of proteins to low pH by having a brief TCA/acetone wash followed by a methanol and acetone wash step, which removes residual TCA and contaminants. An additional methanol wash in the presence of ammonium acetate neutralize residual TCA, increased the pH above 7 which further increased the extraction of proteins by phenol (Wang *et al.*, 2006). The extraction with phenol/SDS buffer have resulted in high quality proteins free from interfering substances as can be seen in the 1-DE as well as the 2-DE gels. The protein loading was also uniform within the samples. The reproducibility of protein extractions were good between extracts since the profiles showed high similarities in terms of protein expression, abundance and banding patterns.

5.17.2 Second Dimension Gel Electrophoresis

Typical 2-DE gel of apple leaf protein revealed that most of the proteins resolved in the pH 4-7 range because most of the proteins fell within that range. The MW of proteins ranged between 10-110 kDa, a pattern observed in leaf tissues such as grape, potato, wheat and sorghum (Aghaei *et al.*, 2008; Basha *et al.*, 2009; Xin *et al.*, 2012; Ngara *et*

al., 2012). The plant proteome consists of a mixture of proteins amongst others membrane proteins, which are hydrophobic. Limitations in proteomics still exclude proteins with extreme *pI* and MW from an analysis (Jain *et al.*, 2008), which were excluded from the IPG strips used during the first phase of protein separation namely isoelectric focusing. These strips with a 4% total acrylamide concentration having a reduced protein uptake may be a further limiting factor in proteomics. The quality of the leaf protein was good due to the protocol used, which took care of contaminants and proteases, thus limiting degradation of proteins. This allowed the proteins to migrate to their respective *pI* in the first phase and MW, second phase. A prominent cluster of proteins around 50 kDa was visible which is evident in most plant proteomes (Kim *et al.*, 2004; Peng *et al.*, 2009; Ngara *et al.*, 2012). These were identified as RuBisCO, which is the most abundant protein in plants (Ellis, 1979). The reproducibility of protein extraction as well as uniform protein loading on gels are crucial factors in comparative proteomics to consider since inconsistencies in these might lead to protein spots to be erroneously regarded as differentially expressed.

5.17.3 Statistical analysis - One-way ANOVA significant protein spots

Using the One-way ANOVA analysis at 95% significance level, the expression pattern of 113 protein spots was shown to be significantly different between the control and the infected apple leave proteomes. However, only 30 spots were selected for MALDI-TOF analysis based on a spot volume fold expression of more than 0.5 (Table 5.4). Spots with a spot volume of less than 0.5 were very faint. Of these 30 spots only 28 were positively identified with only spot numbers 20 and 25 not identified (Table 5.5).

5.17.4 Identification and characterisation of proteins

Second dimension gel electrophoresis only allows for separation and visualisation of proteins but it does not identify proteins. Apple leaf proteins were then identified by a combination of MALDI-TOF MS, MALDI-TOF-TOF MS and database searching 5.13. Access to an organism's genome sequence is vital in linking genes to the proteins (Lin *et al.*, 2003). For apple, the completion of the genome-sequencing project published by Velasco *et al.* (2010), has greatly contributed to the existing apple databases. Therefore until recently, the identification of apple proteins has relied largely on homology searches against sequence information from other green plants as well as apple ESTs.

Out of the 30 spots that were picked for mass spectrometry analysis (spots 1-30; Table 5.4), identities of 28 spots were obtained by matching the MS and MS/MS data to the database and passing the criteria of a q-value = 0.05 (Table 5.4). This gave a 93.3% success rate in protein identification. This high identification success rate of the apple leaf proteins can be attributed to the high number of conserved proteins in higher plants as well as the annotation of the apple genome. Furthermore, the high success rate in the positive identification of apple proteins can be attributed to the use of MALDI-TOF-TOF MS. Tandem mass spectrometry allows for *de novo* sequencing of protein peptides (Dubey and Grover, 2001) and it can play an important role in the identification of proteins of organisms where little or partial genome information exists in databases (Aebersold and Goodlet, 2001).

Of the 30 spots that were picked for MS analysis (Figure 5.4; Table 5.5), 2 were

unidentified (Table 5.5). These unidentified protein spots exhibited varying degrees of protein abundances (Figure 5.4). For this reason, the non-identification of these spots might be due to two main reasons; (i) for some of the low abundant proteins, the protein quantities present in the picked gel plugs might be a limiting factor for positive identification or (ii) their identification might be limited by the unavailability of full annotation of the apple protein database. Spots 20 and 25 might therefore represent apple specific proteins, which are not common amongst other pome fruit or higher plants.

A total of eight proteins were observed to be present in multiple spots within the 2-DE gels e.g., spots nos. 2, 3, 4, 5, 6, 10, 24 and 28 (Table 5.5, Section 5.12.2). The existence of proteins in multiple spots on 2-DE gels has been reported in other proteomic studies of sorghum leaves (Ngara *et al.*, 2012), wheat (Yang *et al.*, 2010) and oilseed rape (Albertin *et al.*, 2009; Yao *et al.*, 2011), amongst others. Multiple spots for a single protein may be due to several reasons such as PTMs, proteolytic degradation of proteins by endogenous proteases, presence of different protein isoforms from the multigene families as well as the chemical modification of proteins during sample preparati on (Albertin *et al.*, 2009; Gao *et al.*, 2009; Yan *et al.*, 2006).

5.17.5 Ontological classification of identified proteins

The compatible interaction between the susceptible apple Golden Delicious cv and the apple scab pathogen *Venturia inaequalis* resulted in the modulation of at least 30 proteins of which only 28 have been positively identified. Thirty protein spots have been identified to be differentially expressed of which 9 proteins showed an increased

abundance, 4 proteins showed a decrease in abundance, 16 absent/silenced and 1 present in the fungal infected samples. Subcellular localisation of proteins defines the function and mechanism of action of the proteins and their interactors for systems biology (Kumar *et al.*, 2002). Proteins were grouped into categories based on their location in the cell or biological process categories according to the annotation in the NCBI nr databases (Figure 5.9). These proteins, however, were also queried using TargetP to determine their location in various cell compartments. As per TargetP, most of the detected proteins were localised in three organelles, the chloroplast (61%), cytoplasm (32%) and the nucleus (7%; Figure 5.9). These results are consistent with the observation that the most prominent functional category amongst the identified apple leaf proteins (Figure 5.4; Table 5.5) was carbohydrate metabolism and photosynthesis related processes (Figure 5.8; Table 5.5). In green plants such as apple, photosynthesis and carbohydrate metabolic pathways occur primarily in photosynthetic organelles called chloroplasts (van Wijk, 2004). Therefore, the dominance of chloroplast as the main subcellular location of apple leaf proteins in this study correlates well with the functional classification results (Table 5.5; Figure 5.9).

Identified proteins were also functionally categorized and many of the proteins related to carbohydrate metabolism/energy (56%), protein destination and storage (10%), nucleotide metabolism and transport (3%), disease and defence (7%), protein synthesis (7%) and finally redox/response to stress (10%).

5.17.5.1 Carbohydrate metabolism/energy

Triose phosphate isomerase

Triose phosphate isomerase (TIM) is a glycolytic enzyme and it is responsible for the catalysis of the reversible interconvention of glyceraldehyde 3-phosphate and di-hydroxyacetone phosphate (Lolis *et al*., 1990). This enzyme plays an important role in a number of metabolic pathways including glycolysis, gluconeogenesis, and Calvin cycle and is essential for efficient energy production (Lolis *et al*., 1990). In this study, this enzyme spots' (spots 2 and 3) showed increased abundance upon the fungal infection. Fang and his group also reported the same observation during the initial stages of strawberry leaves in response to *Colletotrichum fragariae* infection 24 hour post inoculation (Fang *et al*., 2012). It has been previously reported that this enzyme is also involved in stress response. It has also been found that a reduction in TIM expression was reported in plants after abiotic or biotic stress conditions (Riccardi *et al*., 1998; Morris and Djordjevic, 2001). Our finding correlates with a study by Wang *et al.* (2012) where resistant soybean was infected with *Phakopsora pachyrhizi*. In that study, proteins associated with the glycolytic pathway, including TIM, were up regulated after pathogen infection (Wang *et al*., 2012). The speculation is that the up regulation of TIM in this regard could be partly due to increased energy requirements for the detoxification and repair of damage caused by oxidative molecules (Kang *et al*., 2012). The up-regulation of TIM in this study could play a role in detoxification and repair of damaged cells in the plant.

RuBisCO and Sedoheptulose bisphosphatase

The RuBisCO (spots 4 and 24) and sedoheptulose bisphosphatase (spot 23), which are involved in the Calvin cycle have been identified. The Calvin cycle (also termed the reductive pentose phosphate pathway) is a metabolic pathway that produced pentose sugars (Heldt, 1997). The cycle is characterised by three phases; the carboxylation, reduction and regeneration phases (Heldt, 1997). The two proteins identified are involved in two of these phases namely carboxylation and regeneration phases of the Calvin Cycle, respectively. RuBisCO catalyses the carboxylation and oxygenation of ribulose bisphosphate in the Calvin and photorespiratory cycles respectively. RuBisCO is an important catalyst of the Calvin cycle.

RuBisCO is a multimeric enzyme with two subunits; large (50-55 kDa) and small (12-18 kDa) (Andersson and Backlund 2008). In this study both units *viz.* ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (RuBisCO; spots 4 and 24) and ribulose-1,5-bisphosphate carboxylase small unit (spot 8) were identified. However spot 4 showed an increase in abundance, whereas spots 24 and 8 are absent or degraded in the infected apple leave proteome. Marra *et al.* (2010) found in their study, RuBisCO to be up-regulated after infection of *Brassica napus* cv. Surpass 400 with *Leptosphaeria maculans*. Similar findings were found by Subramanian *et al.* (2005) in which an increased level of Calvin cycle enzymes, such as RuBisCO, in the proteome of a resistant *B. napus* cultivar challenged with *L. maculans* were found, while a reduction in photosynthetic activity was observed in a susceptible cultivar. The up-regulation of

RuBisCO (spot 4) is in agreement with the findings of Marra *et al.* (2010) and Subramanian *et al.* (2005).

Sedoheptulose bisphosphatase functions in the regenerative phase of the Calvin cycle, and plays a role in photosynthetic CO_2 assimilation. It catalyses the dephosphorylation of sedoheptulose-1,7-bisphosphate to ribulose-1,5-bisphosphate, which is needed for the initial carboxylation reaction. This reaction is irreversible; therefore the Calvin cycle can proceed only in the direction of pentose phosphates. This enzyme is also an important regulator of the Calvin cycle and therefore affects the rate of photosynthesis directly. However in this study, sedoheptulose bisphosphatase is absent or degraded in the proteome of the infected samples. The absence of RuBisCO small and large unit (spots 8 and 24 respectively) as well as sedoheptulose bisphosphatase (spot 23), which are involved in the regeneration phase of the Calvin cycle, may be indicative that the plant's photosynthetic rate was affected, resulting in a decreased carbon fixation as well as a decreased energy supply.

Photosystem II reaction center

The Photosystem II reaction center consists of a core of intrinsic proteins: D1 and D2 bind the Chl 1, Pheo a, and plastoquinone cofactors that participate in light-catalysed charge separation; the larger chlorophyll-binding polypeptides CP47 and CP43 and cytochrome b559, along with a number of small protein subunits are also associated with the core structure of the photosystem (Popelkova and Yocum, 2011). Three inorganic cofactors (4 Mn, 1 Ca2+, 1 Cl-) and at least three extrinsic proteins are bound to the PSII core proteins, and together they form a module called the oxygen-evolving complex

(OEC) which is the catalytic center for water oxidation (Popelkova and Yocum, 2011). In this study, one of the extrinsic proteins namely PsbP oxygen evolving enhancer protein 2 (spots 6 and 10) were identified. This protein is required for PSII where it increases the affinity of the water oxidation site for Cl and provides the conditions required for high affinity binding of Ca^{2+} (Bandehagh *et al*., 2011). This protein showed decreased abundance for spot 6 and absent or degraded (spot 10) in the proteome of the infected apple leaves. This protein has a molecular weight of 23 kDa (Ifuku *et al*., 2011). The protein in this study has a theoretical molecular weight of 32.43 kDa and an experimental molecular weight of 20 kDa, which could be the same as identified in Ifuku *et al.* (2011).

A second extrinsic protein found in the PS II reaction center *viz.* PsbO (spots 13 and 17) was identified in this study. This protein forms one of the largest extrinsic subunit of the membrane-associated photosynthetic redox enzyme called Photosystem II, which is also called the manganese-stabilising protein. This protein's key role is to ensure maximal efficiency of the water-oxidation reaction. In this study PsbO (spot 13) showed an increase in abundance. The second isoform of this protein namely spot 17 is absent or degraded in the proteome of the infected apple leaves. This protein has a molecular weight of 33 kDa (Ifuku *et al*., 2011) and in this study it has a theoretical molecular weight of 35.06 kDa and an experimental molecular weight of 30 kDa. This protein has a calculated molecular mass of 26.5 kDa based on its amino acid sequence but runs on SDS-PAGE as a 33 kDa protein (Popelkova and Yocum, 2011). This could be as a result of the nature of these proteins to behave as natively unfolded or intrinsically disordered proteins in solution. They are overestimated in size because they have more charge and as

a result they bind poorly to SDS-PAGE, resulting in this anomalous gel migration (Popelkova and Yocum, 2011).

NAD-dependent epimerase/dehydratase

In this study, a membrane protein, namely NAD-dependent epimerase/dehydratase (spot 19) was identified. The NAD-dependent epimerase/dehydratase, whose abundance is reduced following stress treatment, is an enzyme that is highly regulated by ABA (Proietti *et al*., 2013). This protein was found to relate to a Small Ubiquitin-like Modifier (SUMO) conjugating enzyme, which plays a role in sumoylation pathway in *Arabidopsis*. Sumoylation is implicated in ABA-mediated inhibition of root growth (Proietti *et al*., 2013). Furthermore, this family of proteins utilise NAD as a cofactor. The proteins in this family use nucleotide-sugar substrates for a variety of chemical reactions. It contains the NAD(P)-binding domain, which is a commonly found domain with a core Rossmann-type fold. One of the best studied of these proteins is UDP-galactose 4-epimerase which catalyses the conversion of UDP-galactose to UDP-glucose during galactose metabolism.

ATPase AAA core protein

Four protein spots 9, 12, 14 and 15 were identified as ATPase AAA core protein. ATPase AAA core proteins, known as **A**TPases **A**ssociated with diverse cellular **A**ctivities. They form a large protein family, which play a number of roles in the cell including cell-cycle regulation, protein proteolysis and disaggregation, organelle biogenesis and intracellular transport (Kedzierska, 2006). They could also function as molecular chaperones, subunits of proteolytic complexes or independent proteases. Furthermore, they could function as

DNA helicases and transcription factors (Kedzierska, 2006). These proteins are found in all living organisms and have a highly conserved AAA domain known as the AAA module. This AAA module functions in ATP binding and hydrolysis. ATPases in general are membrane bound enzyme complexes. They serve as ion transporters that combine ATP synthesis and hydrolysis with the transport of protons across a membrane (Katam *et al*., 2010). In this study, one of these proteins showed an increase in abundance and the other three isoforms absent or degraded in the proteome of the infected apple leaves.

Phosphoglycolate phosphatase-like

This protein (spot 16) was identified as phosphoglycolate phosphatase-like and showed a decrease in abundance. This protein functions in photosynthesis/photorespiration, which catalyzes the chemical reaction 2-phosphoglycolate and water into glycolate and phosphate.

The findings in the photosynthesis category where proteins were down-regulated is in agreement to Bilgin *et al.* (2010) in which they postulate that biotic damage to foliage caused a near global down-regulation of genes involved in photosynthesis (Bilgin *et al.*, 2010). They found that genes coding for proteins in photosystem I and photosystem II reaction centers, ATP synthase and several elements of the light-harvesting complex associated with PSII were down-regulated by biotic damage (Bilgin *et al.*, 2010). Plants in general respond to stress by slowing their metabolic processes, thus lowering the need for energy generated by photosynthesis (Han *et al.*, 2013). This is a safety mechanism in that if photosynthetic energy levels are not changed accordingly, damaging ROS could be

produced (Han *et al.*, 2013). In this study, a down-regulation and a total silencing of the genes and/or protein expression was observed. This could point to the fact that the fungal infection had different effects at different phases of the pathway. Furthermore, it could point to the plant's strategy to minimize ROS damage.

5.17.5.2 Protein destination and storage

Protein destination

In the category for protein destination and storage, two different proteins namely chaperonin 21 and chaperonin 60 of which two isoforms of the latter were identified. Molecular chaperones play a key role in contributing to cellular homeostasis in cells under favorable and extreme environmental conditions. These proteins aid in protein folding, assembly, translocation and degradation in a series of normal cellular processes. They function in the stabilisation of proteins and membranes as well as assist in protein refolding under stress conditions such as abiotic or biotic stresses. Five major families of HSPs have been reported *viz.* the HSP70 (DnaK) family; the chaperonins (GroEL and HSP60); the HSP90 family; the HSP100 (Clp) family; and the small HSP (sHSPs) family approximately 17- to 30 kDa and ubiquitin (5-8 kDa; Wang *et al.*, 2004). HSPs including HSP100, HSP90, HSP70 and HSP60 are highly conserved proteins, which are up-regulated in cells that have been subjected to elevated temperatures (cold or heat) or various environmental stresses (salinity, drought and oxidative stress) (Maimbo *et al.*, 2007). The nature of these proteins makes it possible for the plant to adapt to sudden changes in the environment and ultimately survive under these lethal conditions (Didelot *et al.*, 2006).

The chaperonin 21 falls in the category of small heat shock proteins. It has been associated with unfolding proteins to form stable complexes and prevent their irreversible aggregation. They could also form co-aggregates with proteins and then efficiently disaggregate (Garofalo *et al.*, 2009). Generally sHSPs have a wide range of cellular functions such as acting as molecular chaperones, decreasing the intracellular level of ROS, thus protecting PSII function during stress (Hajheidari *et al.*, 2005). The function of these small HSPs is unknown and was suggested by Maimbo and his group that it could play a role in the induction of the basal immune response (Maimbo *et al.*, 2007). In this study, these proteins were absent from the fungal infected plants, which could be as a result of the degradation of proteins.

5.17.5.3 Redox/response to stress

In the redox/response to stress category we find the response to stress i.e. heat shock protein 70 (Hsp70), abscisic acid, stress- and ripening-induced (ASR) proteins showed an increase in abundance as well as the redox protein, thioredoxin.

Heat shock protein 70 (Hsp70)

Heat shock protein 70 (Hsp70) is a known stress responsive protein and is induced in response to abiotic stresses such as heat, cold, drought, salinity and cellular stresses. It consists of a highly conserved N-terminal ATPase domain of approximately 44 kDa and a C-terminal peptide-binding domain of approximately 25 kDa (Wang *et al.*, 2004). It prevents the aggregation of stress-denatured proteins and also facilitates the refolding of

proteins to restore their native biological functions (Wang *et al.*, 2004). It is broadly known to function as a molecular chaperone (Miernyk, 1997).

In this study, an increased abundance of HSP70 protein (spot 7) in response to biotic stress was observed. This is expected for this protein to be induced to function in its role as a molecular chaperone to aid in protein refolding of proteins, which tend to aggregate under stress conditions. Wang *et al.* (2004) stated that HSP70 and HSP90 have a role in signal transduction that will ultimately lead to plant defence (Young, 2001; Wang *et al.*, 2004). However, in a study where these two proteins were silenced in *Nicotiana benthamiana,* the induction of hypersensitive response (HR) and also non-host resistance were affected (Kanzaki *et al.*, 2003). Hsp70 proteins only ensure that proteins in stress conditions such as abiotic or biotic are refolded and continue with their native biological functions.

ABA-, stress- and ripening-induced (ASR) proteins

ABA-, stress- and ripening-induced (ASR) proteins are hydrophilic charged low molecular weight plant-specific proteins. The ASR proteins are involved in plant development, senescence and fruit ripening, and respond to ABA and abiotic stresses such as water deficit, salt, cold and limited light (Çakir *et al.*, 2003). These proteins also regulate the transcription of sugar- and abiotic stress-regulated genes in fruit and vegetative tissues. ASR genes have been shown to be up-regulated by ABA and respond to abiotic and biotic stresses (Çakir *et al*., 2003; Liu *et al*., 2010). Moreover, the protein act downstream in the signal transduction pathway for sugar and ABA signals (Çakir *et*

al., 2003). The increase in abundance of the protein in this study is in agreement with the findings of Çakir (2003) in which they found genes expressing ASR proteins were upregulated. This could be indicative of the plant responding to the fungal.

Oxidative burst

Thioredoxin is one of the ROS scavenging enzymes, which play a key role in scavenging ROS and maintain ion homeostasis under abiotic and biotic stresses (Mittler *et al.*, 2004). Thioredoxin (trx) proteins are small (12 kDa) and function as a catalyst in the thiol-disulfide interchange. They are also involved in the regulation of the redox environment of the cell (Gelhaye *et al.*, 2005). This protein also serves as an antioxidant and plays a role in the elimination of reactive oxygen species (ROS), which have damaging effects on the plants. The up-regulation of this protein have been reported under salt stress (Caruso *et al.*, 2008) as well as plants under freeze stress (Han *et al.*, 2013). In this study, an increase in abundance of this protein has been observed.

Plants experience an oxidative burst, which is an excess production of reactive oxygen species (ROS) such as O_2^- and H_2O_2 during pathogen attack (Doke *et al.*, 1996). Reactive oxygen species (ROS) serve as a first line of defence in resistant plants by directly attacking the pathogen (Fang *et al.*, 2012) during the earliest stages of infection by initiating a lipid-peroxidation chain reaction in the plasmalemma (Hajheidari *et al.*, 2005). The ROS induce the production of antioxidants such as thioredoxin to down-regulate ROS production. Thioredoxin in this study showed an increased abundance, an indication that there are ROS present in the cells that needs to be eliminated. This could

also serve as an indication that the plant responded to the invasion by the production of ROS (Doke *et al.*, 1996). This is a natural response of the plant to protect itself against the harmful effects of ROS.

5.17.5.4 Nucleotide metabolism and transport

One protein spot 29, nucleoside diphosphate kinase (NDK) was identified. This NDK is known to be involved in the nucleotide metabolism category and constitute 3% of all the proteins identified. This protein is believed to be a housekeeping enzyme that uses ATP to maintain intracellular levels of CTP for lipid synthesis, UTP for polysaccharide synthesis, GTP for elongation, signal transduction, microtubule polymerisation and provide NTPs for nucleic acid synthesis (Biggs *et al*., 1990). It also plays an important role in hormone responses, heat stress, drought stress, mitogen-activated protein kinase (MAPK)-mediated H_2O_2 signalling, growth and development (Hajheidari *et al*., 2005). In drought stressed sugar beet leaves, NDK was found to be highly up-regulated (Hajheidari *et al.*, 2005). This protein was found to be absent or degraded in the proteome of the infected apple leaves.

5.17.5.5 Protein synthesis

In the category for protein synthesis two proteins were identified namely RraA (spot 27) and IF5A (spot30), which play a role in protein synthesis. Protein synthesis is important in providing the cells with the needed proteins and enzymes, which participate in many biological processes within the cell. The regulator of ribonuclease activity A (RraA)

protein (Spot27) regulates RNA metabolic process. In the degradation of mRNA and the processing of catalytic and structural RNA's in E. coli, ribonuclease E (RNase E) play a key role. It has been found that RraA binds to RNase E and inhibits RNase E endoribonuclease activity, causing global changes in RNA abundance and ultimately alters gene expression, which affects protein synthesis (Lee *et al.*, 2003). In this study RraA was absent or degraded in the proteome of the infected apple leaves, an indication that mRNA can be degraded, which could ultimately influence protein synthesis.

The translation factor eIF5A (spot 30) contains the unusual amino acid hypusine [Nε -(4-amino-2-hydroxybutyl)lysine]. Hypusine is required for eIF5A to associate with ribosomes and to stimulate methionyl-puromycin synthesis, which is a model reaction for the formation of the first peptide bond (Saini *et al.*, 2009). In this way, polypeptides are elongated during synthesis. In this study we found IF5A to be present only in the proteome of the infected apple leaves and absent in the control apple leave proteome. Proteins involved in this category play a role in protein synthesis. Biotic stresses such a fungal infections cause protein damage and/or degradation either due to oxidative damage or proteolytic activities. The absence of spot 27 and the presence of spot 30 only in the infected leaf proteome may function in restoring damaged proteins, which function in restoring the plant's cell metabolic activities and general growth.

5.17.5.6 Disease/defence

In the category of disease/defence two protein spots (spots 5 and 28) were identified as major latex protein (MLP) domain. The MLP protein forms part of the Bet v 1 family,

which is a pathogenesis related (PR10)-like protein family (Chen and Dai, 2010). The Bet v 1 family is divided into 11 subfamilies of which the major latex protein (MLP) is a member (Radauer *et al*., 2008). From literature its molecular weight ranges between 15-18 KDa and a *pI* range of 4.75-6.65. Results in this study are in accordance with literature regarding MW and *pI*. One of the spots (spot 28) is absent, which could be due to degradation. However, spot 5 showed an increase in abundance and does not show any modification of the protein since it falls within the molecular weight and *pI* range for MLP. This protein MLP, a subfamily of *Betula verrucosa* 1 (Betv1) family of proteins, which perform multiple functions such as physiological processes including disease (pathogen invasion) and stress resistance namely abiotic, biotic and development. In this study, an increase in expression of this protein was observed, which could be due to the fungal invasion and could function in stress resistance.

In conclusion, some of the response proteins differentially expressed discussed here does function in priming the host against serious fungal invasion. Differentially expressed proteins identified in this study belong to a number of important functional categories including stress, defence, energy metabolism and protein synthesis. As expected, most of these proteins were identified in other proteomic studies. It is however surprising that no metallothioneins were identified in this project. These proteins were associated with resistance to pathogens in resistant cultivars and were found to be up-regulated (Degenhardt *et al*. 2005). However, since the cultivar in the study is susceptible to the fungus, the reaction with respect to the rate of invasion might be to slow and the fungus

overcomes the resistance. Hence the presence of olivaceous spots at a later stage during the infection.

Chapter Six - Comparative transcriptomic and proteomic profiling of the interaction between *Malus* x *domestica* and *Venturia inaequalis* at Day 4 post inoculation

6 Introduction

In this chapter, results from Chapter five (proteomics) were correlated to results of Day 4 extracted from the transcriptome data set obtained in chapter four. Focus for this chapter was on the proteins and transcripts up-regulated upon *Venturia inaequalis* infection. This chapter would also serve to validate genes differentially up-regulated at Day 4 post inoculation. The actual presence of proteins differentially up-regulated in the proteome of Day 4 would be a validation thereof.

6.1 Extraction and identification of DE genes for Day 4

Command-line scripts and in-house scripts (Appendix B) were used to extract the differentially expressed genes from the pool of differentially expressed genes generated in Table 4.2 (section 4.5.1) for Day4. Extracted transcripts were mapped to the apple genome to extract the corresponding gene from the reference genome (Appendix B) Sequences were then submitted to Blast2GO® version 2.7.0 (Conesa *et al*., 2005) for functional annotation.

6.2 Functional annotation

The effect of the treatment as discussed in 4.5.1 resulted in 814 differentially expressed (DE) genes of which 162 were up-regulated and 652 were down-regulated. The 162 genes mapped to 395 transcripts in the apple gene-finding format (gff3) file (http://www.rosaceae.org/species/malus/malus_x_domestica/genome_v1.0). The transcript sequences obtained from the apple genome as discussed in 4.5.2 were functionally annotated by loading them on Blast2GO® version 2.7.0 (Conesa *et al*., 2005), for BlastX and gene ontology (GO) analysis, as discussed in detail in section 2.14.2.6. Of these, 327 found a BLAST hit and quite a number of these had one or more GO terms assigned to it. However, the focus of discussion will be on the up-regulated genes. Proteins encoded by the up-regulated transcripts were functionally categorised (Figure 6.1, Table 6.1 – Appendix A). Categories affected by the up-regulated genes are signal perception (21%), defence (3%), transport (10%), detoxification (24%), disease resistant (10%), transcription (17%), hormone synthesis (3%), protein synthesis (3%), protein destination and storage (3%) and response to abiotic stress (3%).

In the infected apple leaf proteome of Day 4, 30 differentially expressed proteins were selected of which 28 were identified. From these 9 were up-regulated upon fungal infection Figure 6.2 (Table 6.2 –Appendix A). These were involved in many categories including energy (56%), disease/defence (11%) and redox/response to stress (33%).

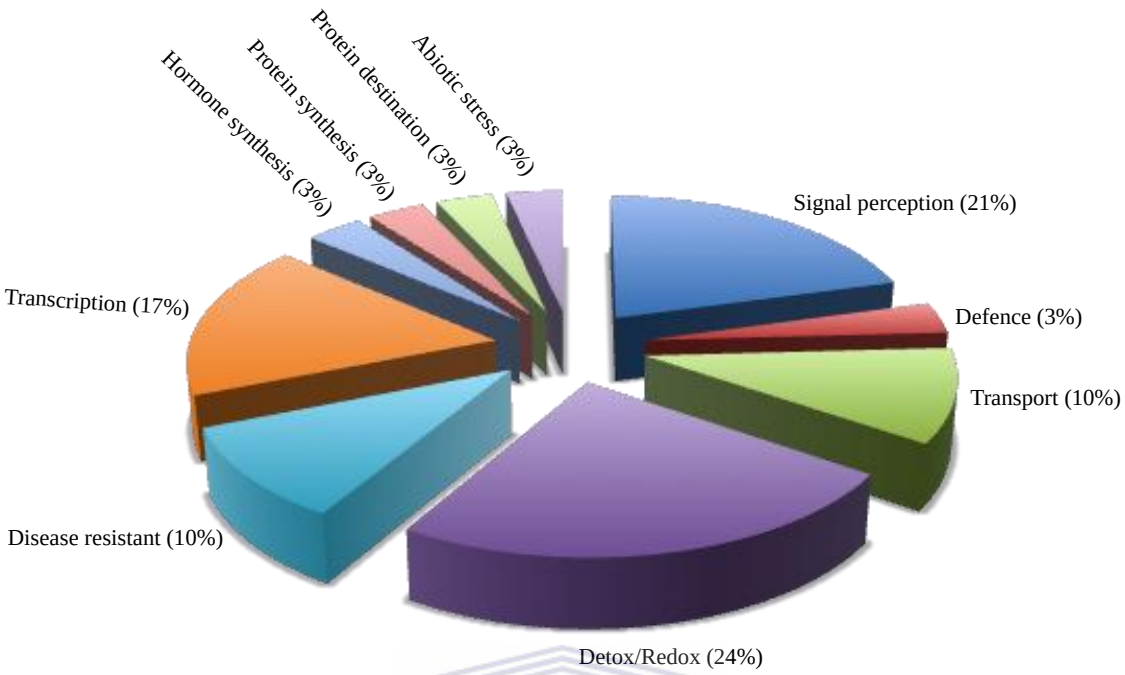


Figure 6.1 Transcripts up-regulated at 4 dpi. Functional categories affected during the infection.

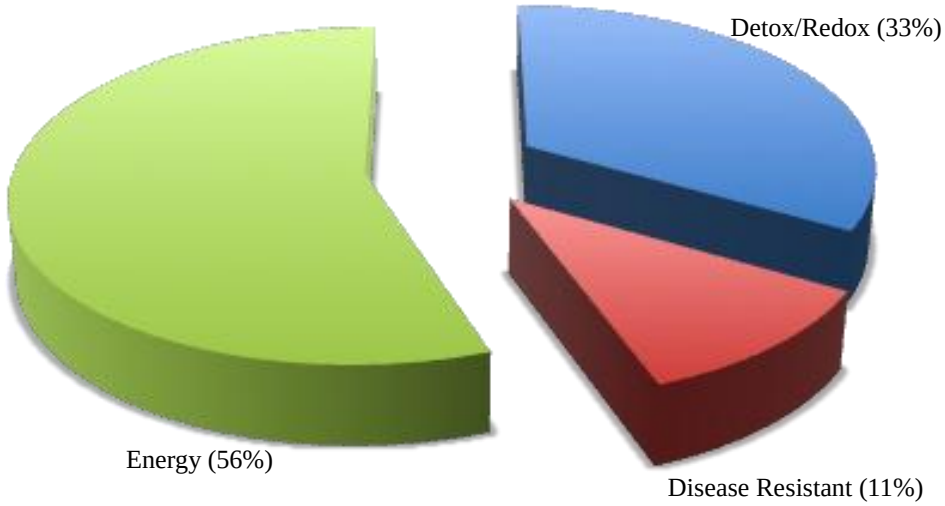


Figure 6.2 Proteins that showed an increase in abundance upon infection at 4 dpi. Functional categories affected during the infection.

6.3 Discussion

RNA-Seq data obtained from the sequencing of cDNA libraries for the five time points in chapter four allowed the extraction of transcripts for Day 4 versus Day 0. Access to the apple genome sequence data made annotation of the genes possible. The newly developed pipeline in chapter four was used to analyse the data set with the view to identify responsive proteins produced at a very early stage of infection with *Venturia inaequalis*. The results were integrated with data generated from proteomics analysis performed in Chapter five and in previous reports (Gau *et al.*, 2004) to build a picture of how the plant respond to an invading fungus with particular emphasis on pathogenesis related-, disease resistant-, cell wall modification- and signal perception proteins.

Transcriptome analysis of the Day 4 infected transcriptome identified 162 genes, which mapped to 395 apple transcripts. Of these, 327 found positive protein identification. Included in these proteins (Table 6.1) were those involved signal perception i.e. RLKs, resistance proteins such as LRR-RLKs, transcription factors of interest were wrky70, ring and zinc finger; hormone synthesis (gibberellin 3-beta-dioxygenase 4-like); stress/detoxification (thioredoxin, cytochrome p450); defence related proteins such as PR-proteins glucan endo beta-glucosidase, disease resistance proteins such as disease resistant protein and syntaxin, transport such as ATP-binding cassette (ABC) transporters and secondary metabolites.

In the transcriptome analysis, genes up-regulated, affected the same categories (Fig. 6.1) of defence mechanisms that were induced in chapter 4. Firstly signal perception followed

by transcription factors. The signal perception proteins are part of the plasma membrane and are known as integral plasma membrane proteins. These proteins include the RLK's such as LRR, serine threonine-protein kinase, LRR-RLKs, which is one of the largest family of receptor kinase proteins (Yadeta *et al.*, 2013). They are quite abundant in plasma membrane. It is not surprising to see that the transcript levels for the genes encoding for these proteins are up-regulated since they form part of the basal-defence system. The signal perception proteins were absent from the proteome analysis.

Furthermore, defence related proteins such as glucan endo-β-glucosidase 11 were up-regulated whereas the pathogenesis related proteins (PR) such as PR-1, PR-4, PR-5 and PR 10, which were present in the transcriptome analysis chapter, were absent at this stage. The glucan endo-β-glucosidase 11 may act in protecting the plant against fungal infection through its ability to degrading fungal cell wall polysaccharides (Xu *et al.*, 1992). However, none of the PR-proteins were present in this proteome analysis but it was induced in the study of Gau *et al.* (2004).

Disease resistance protein and syntaxin were also induced upon infection. Disease resistant proteins were discussed in chapter 4. Syntaxin in plants is a disease resistant protein, which is involved in gene-for-gene, basal and SA-associated defence in plants against pathogens (Kalde *et al.*, 2007).

In this study, no cell wall modification genes were up-regulated. Genes involved in lignin synthesis such as phenylalanine ammonia lyase (PAL), lacto-peroxidase and peroxidase

were absent from this data set. This was also absent from the proteomics data. It seems that early in the infection the plant has only sensed the intruder and has not yet started with cell wall modifications yet. It is however induced later on during the infection stage as discussed in chapter 4.

The proteomics analysis was consistent to some extent in that proteins up-regulated influenced the same functional categories (Fig. 6.2) as those for the transcriptome. Categories affected included energy, redox/response to stress, detoxification and disease/defence.

In the energy category, genes encoded were involved in photosynthesis. This increase in energy demand could probably play a role in the detoxification of cells and repair of damage inflicted by the oxidative molecules. An up-regulation of proteins involved in the redox/response to stress category was observed. These proteins are ABA stress- and ripening (ASR), HSP70 and thioredoxin. ASR proteins were induced in reaction to stress responses such as abiotic and biotic stresses. The HSP70 protein is involved in protein stabilising after biotic or abiotic stress. Thioredoxin serves as an antioxidant and plays a role in eliminating reactive oxygen species from the cells. A thioredoxin-like 1-chloroplastic protein has been found in the transcription data to be up-regulated.

In summary, the RNA-Seq analysis of the infected apple leaf transcriptome was successful in identifying the response proteins produced upon infection with the fungus. Integration of this analysis with corresponding Day 4 proteomics data is important for the study of Apple-*Venturia inaequalis* host-pathogen relationships to distinguish proteins

that are produced early during the infection and those produced at a later stage of infection. Additionally this integrated approach identified major signal, transcription factor, defence, disease resistant and detoxification proteins that may be targeted for future research.

Chapter Seven - General Discussion and Conclusion

7 Introduction

In this chapter the main findings of the study are discussed. Each chapter was discussed in detail within the document.

The aim of this project was to analyse the interaction between *Malus* x *domestic* and *Venturia inaequalis*, the causal agent of apple scab on apples. In order to answer the research question, proteomics and transcriptomics were used as tools to study this interaction at a molecular level. The recently sequenced apple genome (Velasco *et al*., 2010) paved the way to conduct this research using these tools. The results generated in this study would provide insight on how the apple proteome and transcriptome of a susceptible apple cultivar changed in response to the fungal infection. These results could then be used to complement research done on similar projects in other apple cultivars. This would also add knowledge to the current apple intellectual database.

7.1 Molecular characterisation of *Venturia inaequalis*

In this study the characterisation of the fungal strain was of utmost importance since the credibility of the project was based on the identity of the *Venturia inaequalis* strain. Previous studies performed by other groups such as Broggini *et al.* (2010), Gladieux *et al.* (2008) and Guerin *et al.* (2007) clustered different isolates of *Venturia inaequalis* into

their respective groups of origin. Guérin *et al.* (2004, 2007) designed 21 polymorphic microsatellite markers, which, were successfully used in their study to characterise and to cluster *Venturia inaequalis*. Based on these studies, an attempt was made to characterise a local South African strain obtained from Golden Delicious infected apple leaves. All the primers used in this study amplified alleles, which correspond to those amplified in the studies of Broggini *et al.* (2010), Gladieux *et al.* (2008) and Guerin *et al.* (2007). Further support for this finding came from the study of Celton *et al.* (2010) in which they have used a draft sequence genome from the same local strain as a reference genome in anchoring contigs to SNPs.

Thus, the 21 polymorphic microsatellite markers used in this study could confirm the strain used to be *Venturia inaequalis*. These markers could serve as a diagnostic set for future strain or isolate characterisation.

7.2 Transcriptome analysis

RNA-Seq technology has revealed the intricate network of genes involved in the interaction between the apple host and its pathogen, *Venturia inaequalis*. It is clear from the results obtained that the host reacted by preparing against a fungal attack. Firstly, the plant reacted as expected by activation of its surveillance system *viz.* detection of fungal spores on the leaf. A number of signalling genes have been induced to alert the plant of a possible intruder. These are the RKL's such as the leucine rich kinase and the cysteine rich kinase receptors. These genes are activated to prime the plant for the fungal attack.

This was followed by the activation of the transcription factors such as wrky, myb10 and zinc fingers to drive the processes.

Structurally plants adapt or make changes in its cell wall to combat the fungus. This was achieved through the lignification of the cell wall to prevent the fungus from spreading through to the rest of the leaf. This also causes the fungus to starve as a result of this cell wall fortification because nutrients are deprived from the fungus. Genes in the SA pathway have been induced which were an indication that the plant combats the fungus on a hormonal level. This was in accordance to literature in which plants react to biotrophic fungal invasion by activating the SA pathway (Felton *et al.*, 1999a,b; Koorneef and Pieterse, 2008). In a necrotrophic fungal invasion the JA and ethylene pathways are activated to combat these kinds of fungi (Felton *et al.*, 1999a,b; Koorneef and Pieterse, 2008).

Furthermore, the chemical defence of the host have been activated. A number of R genes have been induced to fight the fungal invasion. Amongst these are the NBS-LRR, disease resistance proteins RPM1, disease resistance RGA3, enhanced disease susceptibility 1 (EDS1). The EDS1 gene has also been implicated in SA biosynthesis (Pieterse and Van Loon, 2004; Wiermer *et al.*, 2005) These genes were induced in accordance to the gene-for-gene relationship in order to combat the fungal attack.

An increase in expression of quite a number of PR genes such as the PR-1, PR-2, PR-5 and PR-9 were observed. This is in agreement with literature, which shows that the

activation of these genes is markers for SA response. This has been shown in *Arabidopsis* infected with cabbage leaf curl virus (CaLCuV) and were elevated (Ascencio-Ibáñez *et al.*, 2008).

The limitation of this transcriptome analysis is the absence of replicates since cost was a limiting factor considering the scale of the project. As a result of this, progression of a disease resistance gene for example, could not be traced. To address this limitation, replicates had to be included in the analysis, which would have given results that were unbiased.

7.3 Proteome analysis

In the proteome analysis, Day 4 uninfected and infected apple leaf proteome was qualitatively and quantitatively analysed. Quantitative spot analysis revealed 30 spots to be differentially expressed. Of these 30 spots, 28 were positively identified through homology searches against the non-redundant NCBI database and manually annotated by using the apple protein database. Through this most of the positive identifications obtained through NCBI, were assigned positive apple protein identification. These proteins were functionally categorised into different categories such as energy, redox/response to stress, protein destination and storage, nucleotide metabolism and transport, protein synthesis and disease/defence. A number of multiple spots for the same protein were observed, which could be isoforms of the same protein. Some protein spots were present in the control leaf proteome and absent in the infected leaf proteome or visa versa. Proteins differentially expressed were in response to the fungal infection and prepared the plant for fungal invasion. At this point, proteins involved in the

carbohydrate metabolism were up-regulated and/or absent in the infected apple leaf proteome. Furthermore, proteins involved in damage control caused by oxidative molecules and a subsequent detoxification of cells were up-regulated. Proteins such as the chaperonins and in particular heat shock 70 protein were up-regulated as a result of the biotic stress in order to stabilise the proteins.

Furthermore, proteins identified to be involved in nucleotide metabolism and protein synthesis were present and/or absent in the infected apple leaf proteome. Proteins identified as pathogenesis related (PR) proteins were up-regulated. In conclusion, proteomics as a tool can be successfully applied in the identification of responsive proteins in the apple and *Venturia inaequalis* interactions. This can be supported with the numerous publications in this field (Kim *et al.*, 2003; Gau *et al.*, 2004; Liang *et al.*, 2008; Lin *et al.*, 2008; Yu *et al.*, 2008; Ryu *et al.*, 2009; Cao *et al.* 2009; Huang *et al.*, 2009; Marra *et al.*, 2010; Murphy *et al.* 2010; Nanjo *et al.*, 2010; Sobhanian *et al.*, 2010; Shin *et al.*, 2011; Wang *et al.*, 2012; Li *et al.*, 2013).

7.4 Integration of proteomic and transcriptomic data

In this chapter, the proteome data obtained from chapter four was integrated with the transcriptome data set of Day 4. However, the focus was on the up-regulated data of both analyses. In the proteome data, proteins, which were differentially up-regulated, affected the same categories as those, which were affected in the transcriptome data set of Day4. Only one protein i.e. thioredoxin in the proteome and a close homolog to thioredoxin in the transcriptome were up-regulated. Although only one protein correlates for both data

sets, it is worth mentioning that at a broader scale, the genes induced in these affect the same functional categories. The up-regulation of genes affecting the signal perception, transcription, stress/detoxification, defence, disease resistance proteins and transport categories were observed. The response of the host is in agreement with literature where genes in these categories were up-regulated (Pritsch *et al.*, 2000; Campos-Soriano *et al.*, 2010; Venu *et al.*, 2010; Zhang *et al.*, 2011; Xin *et al.*, 2012; Gusberti *et al.*, 2013).

From the data, the conclusion was drawn, that the level of quite a number of transcripts were increased which encode for proteins to be up-regulated, however only a few hundred proteins spots were visible. This could be attributed to the limitations of the 2-DE gel used. Most of the proteins up-regulated in the transcriptome data set were integral plasma membrane proteins which are often underrepresented in gel-based analyses, due to their physiochemical heterogeneity, hydrophobicity, and low abundance (Marmagne *et al.*, 2007). Furthermore, they have a high molecular weight and of alkaline in nature (Gilmore and Washburn, 2010). The gel-based proteomics technique however suffers drawbacks, since buffers used in the second dimension gel exclude hydrophobic and high molecular weight proteins from gels (Jain *et al.*, 2008). Membrane proteins in this proteome analysis were identified, which could have resolved as a result of the chaotropes and detergents used in solubilising the proteins for 1-DE and 2-DE gels (Yadeta *et al.*, 2013). To address these limitations, whole-proteome analysis could be implemented by a high-throughput liquid chromatography approach, i.e. a liquid chromatography-tandem MS (LC-MS/MS) for large-scale identification and

quantification of proteins in particular membrane proteins. This approach is ideal for separating peptides from complex protein digests.

7.5 Future work

The realisation of early inception without the use of fungicides would be ideal in order to disable the fungus. In order to understand the pathogenesis of the fungus, an *in planta* fungal gene expression analysis should be performed. These can be achieved through robust transcript profiling methods such as RNA-Seq (Mortazavi *et al*., 2008). Extracting total RNA from bulk plant tissue for profiling suffer drawbacks such as low signal/background ratio (Tang *et al*., 2012). It is therefore difficult to separate pathogen infection structures from surrounding plant tissue after penetration. However techniques such as laser capture micro-dissection (LCM) has become an important tool for isolating individual cells or cell types (Nelson *et al*., 2006). This technique allows for the isolation of both fungal and host plant cells after pathogen infection for molecular studies. Furthermore transgenic fungal strains that constitutively express fluorescent proteins such as green fluorescent protein (GFP) (Maor *et al*., 1998) and AmCyan (Bourett *et al*., 2002) in the fungal cytosol to visualise the infection structures, has made it possible to harvest specific fungal cells during infection of plant tissues. These techniques can be used to provide relatively uniform, pathogen-rich samples for downstream genome-wide gene profiling via RNA-Seq. The advantage of using LCM is, it allows for maximising signal-to-noise ratio for *in planta* fungal gene expression and facilitate developmentally staged deep profiling of fungal genes during infection (Tang *et al*., 2006).

Sample preparation is a crucial step in LCM to ensure the isolation of contaminated-free and RNA with high integrity (Tang *et al*., 2006). In a method used by Tang *et al.* (2006) samples were fixed with appropriate fixatives to preserve the integrity of cell morphology and target metabolites (e.g. RNA), embedding the fixed tissue in paraffin wax for sectioning onto microscope slides. The sample sections are then deparaffinised, rehydrated, and cells are dissected using a laser, focused through a microscope. LCM samples are collected into protective (e.g. RNAse-free) medium for isolation of RNA. The RNA can then be subjected to gene expression studies such as RNA-Seq (Mortazavi *et al*., 2008), which is useful in studying or analysing of the expression levels for mRNAs in a given cell or tissue for any given time. The use of bioinformatics tools will enable us to mine through the data to find differentially expressed genes, which are involved in the early infection process. Differentially expressed genes identified can now be targeted to disable the fungus at a very early stage.

Reference:

Aarts, N., Metz, M., Holub, E., Staskawicz, B. J., Daniels, M. J., & Parker, J. E. (1998). Different requirements for *EDS1* and *NDR1* by disease resistance genes define at least two R gene-mediated signaling pathways in *Arabidopsis*. *Proceedings of the National Academy of Sciences, 95*, 10306–10311.

Acharya, B. R., Raina, S., Maqbool, S.B. Jagadeeswaran, G., Mosher, S. L., Appel, H. M., Schultz, J. C., Raina, R. (2007). Overexpression of *CRK13*, an *Arabidopsis* cysteine-rich receptor-like kinase, results in enhanced resistance to *Pseudomonas syringae. Plant Journal*, *50*, 488–499.

Adams, M. D., Kelley, J. M., Gocayne, J. D., Dubnick, M., Polymeropoulos, M. H., Xiao, H., Venter, J. G. (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. *Science, 252*, 1651–1656.

Aebersold, R., & Goodlet, D. R. (2001). Mass spectrometry in proteomics. *Chemical Reviews*, *101*, 269–295.

Aghaei, K., Ali, A. E., & Setsuko, K. (2008). Proteome analysis of potato under salt stress. *Journal of Proteome Research*, *7*, 4858–4868.

Al-Samarrai, T. H., & Schmid, J. (2000). A simple method for the extraction of fungal genomic DNA. *Letters in Applied Microbiology*, *30*, 53–56.

Albertin, W., Langella, O., Joets, J., Negroni, L., Zivy, M., Damerval, C., & Thiellement, H. (2009). Comparative proteomics of leaf, stem, and root tissues of synthetic *Brassica napus*. *Proteomics*, *9*, 793–799.

Alwine, J. C., Kemp, D. J., & Stark, G. R. (1977). Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with

DNA probes. *Proceedings of the National Academy of Sciences of the United States of America*, *74*, 5350–5354.

An, S. H., Sohn, K. H., Choi, H. W., Hwang, I. S., Lee, S. C., & Hwang, B. (2008). Pepper pectin methylesterase inhibitor protein CaPMEI1 is required for antifungal activity, basal disease resistance and abiotic stress tolerance. *Planta*, *228*, 61–78.

Anders, S. (2011). Analysis of high-throughput sequencing data with Python, http://www-huber.embl.de/users/anders/HTSeq/. Date accessed 30 October 2013.

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biology*, *11*, 106-112.

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, *22*, 2008–2017.

Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. *Plant Physiology and Biochemistry*, *46*, 275–291.

Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T-J., Chu, T-M., Wolfinger, R. D., Cella, R., & Hanley-Bowdoin, L. (2008). Global analysis of *Arabidopsis* gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiology*, *148*, 436–454.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Sherlock, G. (2000). Gene Ontology : tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Au, K. F., Jiang, H., Lin, L., Xing, Y., & Wong, W. H. (2010). Detection of splice junctions from paired-end RNA-seq data by SpliceMap. *Nucleic Acids Research*, *38*, 4570–4578.

Babraham Bioinformatics. (2012). FASTQC, a quality control tool for the high throughput sequence data. *Http://www.bioinformatics.bbsrc.ac.uk/projects/fastq/*. Date accessed 10 January 2013.

Backman, P. (1978). Fungicide formulation: relationship to biological activity. *Phytopathology*, *16*, 211–237.

Baginsky, S., Hennig, L., Zimmermann, P., & Gruissem, W. (2010). Gene expression analysis, proteomics, and network discovery. *Plant Physiology*, *152*, 402–410.

Bainbridge, M. N., Warren, R. L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Jones, S. J. M. (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. *Biomed Central Genomics*, *7*, 246-256.

Bandehagh, A., Salekdeh, G. H., & Toorchi, M. (2011). Comparative proteomic analysis of canola leaves under salinity stress. *Proteomics*, *11*, 1965–1975.

Basha, S. M., Katam, R., Vasanthaiah, H., & Matta, F. (2009). Proteome analysis of muscadine grape leaves. *International Journal of Wine Research*, *1*, 1–13.

Becker-Andre, M., & Hahlbrock, K. (1989). Absolute mRNA quantification using the polymerase chain reaction (PCR). A novel approach by a PCR aided transcript titration assay (PATTY). *Nucleic Acids Research*, *17*, 9437–9446.

Benaouf, G., & Parisi, L. (2000). Genetics of host-pathogen relationships between *Venturia inaequalis* races 6 and 7 and *Malus* species. *Phytopathology*, *90*, 236–242.

Benjamini, Y., & Hochberg, Y. (2000). On the adaptive control of the false discovery fate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, *25*, 60–83.

Benyagoub, M., Benhamou, N., & Carisse, O. (1998). Cytochemical Investigation of the Antagonistic Interaction Between a *Microsphaeropsis sp*. (Isolate P130A) and *Venturia inaequalis*. *Phytopathology, 88*, 605–613.

Berardini, T., Mundodi, S., Reiser, L., Huala, E., Garcia-Hernandez, M., Zhang, P., Rhee, S. (2004). Functional annotation of the *Arabidopsis* genome using controlled vocabularies. *Plant Physiology, 135*, 745–755.

Berggren, K., Chernokalskaya, E., Steinberg, T. H., Kemper, C., Lopez, M. F., Diwu, Z., Patton, W. F. (2000). Background-free, high sensitivity staining of proteins in one- and two dimensional sodium dodecyl sulfate-polyacrylamide gels using a luminescent ruthenium complex. *Electrophoresis*, *21*, 2509–2521.

Bevan, M., Bancroft, I., Bent, E., Love, K., Goodman, H., Dean, C., Chalwatzis, N. (1998). Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. *Nature*, *39*, 485-488.

Bhadauria, V., Banniza, S., Wang, L. X., Wei, Y. D., & Peng, Y. L. (2010). Proteomic studies of phytopathogenic fungi, oomycetes and their interactions with hosts. *European Journal of Plant Pathology*, *126*, 81–95.

Biggs, A. R. (1997). Infectious diseases (Part I): diseases caused by fungi: foliar and friut dieases. *In: Jones AL and Aldwinkle HS. (eds) Compendium of Apple and Pear Diseases. APS Press, St. Paul, Minnesota, pp 6-9.*

Biggs, J., Hersperger, E., Steeg, P. S., Liotta, L. A., & Shearn, A. (1990). A *Drosophila* gene that is homologous to a mammalian gene associated with tumor metastasis codes for a nucleoside diphosphate kinase. *Cell*, *63*, 933–940.

Bilgin, D. D., Zavala, J. a, Zhu, J., Clough, S. J., Ort, D. R., & DeLucia, E. H. (2010). Biotic stress globally downregulates photosynthesis genes. *Plant, Cell and Environment*, *33*, 1597–1613.

Blaedow, R., Chaney, W., Pecknold, P., & Holt, H. (2006). Investigation of fungicidal properties of the tree growth regulator paclobutrazol to control apple scab. *Journal of Arboric*, *32*, 67–73.

Bolwell, G. P., Bindschedler, L. V., Blee, K. A., Butt, V. S., Davies, D. R., Gardner, S. L., Minibayeva, F. (2002). The apoplastic oxidative burst in response to biotic stress in plants: a three-component system. *Journal of Experimental Botany*, *53*, 1367–1376.

Bourett, T. M., Sweigard, J. A., Czymmek, K. J., Carroll, A., & Howard, R. J. (2002). Reef coral fluorescent proteins for visualizing fungal pathogens. *Fungal Genetics and Biology*, *37*, 211–220.

Bowen, J. K., Mesarich, C. H., Bus, V. G. M., Beresford, R. M., Plummer, K. M., & Templeton, M. D. (2011). *Venturia inaequalis*: the causal agent of apple scab. *Molecular Plant Pathology*, *12*, 105–122.

Bradford, M. (1976). A rapid and sensitive method for quantification of micro-gram quantities of protein utilizing the principle of protein-dye-binding. *Analytical Biochemistry*, *72*, 248–254.

Broggini, G. A. L., Bus, V. G. M., Parravicini, G., Kumar, S., Groenwold, R., & Gessler, C. (2010). Genetic mapping of 14 avirulence genes in an EU-B04 X 1639 progeny of *Venturia inaequalis*. *Fungal Genetics and Biology 10*, 1–11.

Brutus, A., Sicilia, F., Macone, A., Cervone, F., & De Lorenzo, G. (2010). A domain swap approach reveals a role of the plant wall-associated kinase 1 (*WAK1*) as a receptor of oligogalacturonides. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 9452–9457.

Burchill, R. T., Hutton, K. E., Crosse, J. E., & Garrett, C. M. E. (1965). Inhibition of the perfect stage of *Venturia inaequalis* (Cooke) Wint., by urea. *Nature*, *205*, 520–521.

Bus, V. G. M., Laurens, F. N. D., van de Weg, W. E., Rusholme, R. L., Rikkerink, E. H. A., Gardiner, S. E., Plummer, K. M. (2005). The *Vh8* locus of a new gene-for-gene interaction between *Venturia inaequalis* and the wild apple *Malus sieversii* is closely linked to the *Vh2* locus in *Malus pumila* R12740-7A. *New Phytologist*, *166*, 1035–1049.

Bus, V., Rikkerink, E., Aldwinckle, H. S., Caffier, V., Durel, C. E., Gardiner, S., van de Weg, W. E. (2009). A Proposal for the Nomenclature of *Venturia inaequalis* Races. *Acta Horticulturae*, *814*, 739–746.

Çakir, B., Agasse, A., Gaillard, C., Saumonneau, A., Delrot, S., & Atanassova, R. (2003). A Grape ASR Protein Involved in Sugar and Abscisic Acid Signaling. *The Plant Cell*, *15*, 2165–2180.

Campos-Soriano, L., Garcia-Garrido, J. M., & San Segundo, B. (2010). Activation of basal defense mechanisms of rice plants by *Glomus intraradices* does not affect the arbuscular mycorrhizal symbiosis. *New Phytologist*, *188*, 597–614.

Cao, T., Kim, Y. M., Kav, N. N. V, & Strelkov, S. E. (2009). A proteomic evaluation of *Pyrenophora tritici-repentis*, causal agent of tan spot of wheat, reveals major differences between virulent and avirulent isolates. *Proteomics*, *9*, 1177–1196.

Carli, M. D, Zamboni, A., Pe, M. E., Pezzotti, M., Lilley, K. S., Benvenuto, E., Desiderio, A. (2011). Two-Dimensional Differential in Gel Electrophoresis (2D-DIGE) Analysis of Grape Berry Proteome during Postharvest Withering research articles, *Journal of Proteome research, 10*, 429–446.

Caruso, G., Cavaliere, C., Guarino, C., Gubbiotti, R., Foglia, P., & Lagana, A. (2008). Identification of changes in *Triticum durum* L. leaf proteome in response to salt stress by two-dimensional electrophoresis and MALDI-TOF mass spectrometry. *Analytical and Bioanalytical Chemistry*, *391*, 381–390.

Celton, J-M., Christoffels, A., Sargent, D. J., Xu, X., & Rees, D. J. G. (2010). Genome-wide SNP identification by high- throughput sequencing and selective mapping allows sequence assembly positioning using a framework genetic linkage map. *Biomed Central Biology*, *8*, 155-162.

Chen, J-Y., & Dai, X-F. (2010). Cloning and characterization of the *Gossypium hirsutum* major latex protein gene and functional analysis in *Arabidopsis thaliana*. *Planta*, *231*, 861–873.

Chen, K., Du, L., & Chen, Z. (2003). Sensitation of defense responses and activation of programmed cell death by a pathogen-induced receptor-like protein kinase in *Arabidopsis*. *Plant Molecular Biology*, *53*, 61–74.

Chen, K. G., Fan, B. F., Du, L. Q., & Chen, Z. (2004). Activation of hypersensitive cell death by pathogen induced receptor-like protein kinases from *Arabidopsis*. *Plant Molecular Biology*, *56*, 271–283.

Chen, X. (2011). Innate immunity in rice. *Trends in Plant Science*, *16*, 451–459.

Cherk, C., Somerville, S., He, K., & Hématy, K. (2009). Host-pathogen warfare at the plant cell wall. *Current Opinion in Plant Biology*, *12*, 406–413.

Chevalier, M., Lespinasse, Y., & Renaudin, S. (1991). A microscopic study of the different classes of symptoms coded by the *Vf* gene in apple for resistance to scab (*Venturia inaequalis*). *Plant Pathology*, *40*, 249–256.

Chevallet, M., Luche, S., & Rabilloud, T. (2006). Silver staining of proteins in polyacrylamide gels. *Nature Protocols*, *1*, 1852–1858.

Chi, F., Yang, P., Han, F., Jing, Y., & Shen, S. (2010). Proteomic analysis of rice seedlings infected by *Sinorhizobium meliloti*. *Proteomics*, *10*, 1861–1874.

Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., & Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*, *21*, 3674–3676.

Conesa, A., & Stefan, G. (2008). Blast2GO : A Comprehensive Suite for Functional Analysis in plant genomics. *International Journal of Plant Genomics*, *2008*, 1-12.

Cornille, A., Gladieux, P., Smulders, M. J. M., Rolda, I., Cam, B. Le, Nersesyan, A., Giraud, T. (2012). New Insight into the History of Domesticated Apple : Secondary Contribution of the European Wild Apple to the Genome of Cultivated Varieties, *PLOS Genetics*, *8*, 1-13.

Cova, V., Paris, R., Passerotti, S., Zini, E., Gessler, C., Pertot, I., Komjanc, M. (2009). Mapping and functional analysis of four apple receptor-like protein kinases related to LRPKm1 in *HcrVf2*-transgenic and wild-type apple plants. *Tree Genetics and Genomes*, *6*, 389–403.

Crosby, J. A., Janick, J., Pecknold, P., Korban, S. S., O'Connor, P., Ries, S., Voordeckers, S. (1992). Breeding apples for scab resistance: 1945-1990. *Acta Horticulturae*, *314*, 43–70.

Culley, D. E., Horovitz, D., & Hadwiger, L. (1995). Molecular characterization of disease-resistance response gene *DRR206-d* from *Pisum sativum* (L.). *Plant Physiology*, *107*, 301–2.

Czernic, P., Visser, B., Sun, W. N., Savoure, A., Deslandes, L., Marco, Y., Verbruggen, N. (1999). Characterization of an *Arabidopsis thaliana* receptor-like protein kinase gene activated by oxidative stress and pathogen attack. *The Plant Journal*, *18*, 321–327.

Davin, L., & Lewis, N. G. (2000). Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis. *Plant Physiology*, *123*, 453–462.

De Cremer, K., Mathys, J., Vos, C., Froenicke, L., Michelmore, R. W., Cammue, B. P., & De Coninck, B. (2013). RNAseq-based transcriptome analysis of *Lactuca sativa* infected by the fungal necrotroph *Botrytis cinerea*. *Plant, Cell and Environment*, *36*, 1992–2007.

Degenhardt, J., Al-Masri, A. N., Kürkcüoglu, S., Szankowski, I., & Gau, A. E. (2005). Characterization by suppression subtractive hybridization of transcripts that are differentially expressed in leaves of apple scab-resistant and susceptible cultivars of *Malus* x *domestica*. *Molecular Genetics and Genomics*, *273*, 326–335.

Dennis, G, J., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, A. (2003). DAVID : Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, *4*, 60-65.

Didelot, C., Schmitt, E., Brunet, M., Maingret, L., Parcellier, A., & Garrido, C. (2006). Heat shock proteins: endogenous modulators of apoptotic cell death. *Handbook of Experimental Pharmacology*, *172*, 171–198.

Doke, N., Miura, Y., Sanchez, L. M., Noritake, T., Yoshioka, H., & Kawakita, K. (1996). The oxidative burst protects plants against pathogen attack: mechanism and role as an emergency signal for plant bio-defence- a review. *Gene*, *179*, 45–51.

Dubey, H., & Grover, A. (2001). Current initiatives in proteomics research: the plant perspective. *Current Science*, *80*, 262–269.

Dubois, E., Fertin, M., Burdese, J., Amouyel, P., Bauters, C., & Pinet, F. (2011). Cardiovascular proteomics: translational studies to develop novel biomarkers in heart failure and left ventricular remodeling. *Proteomics. Clinical Applications*, *5*, 57–66.

Eichhorn, H., Klinghammer, M., Becht, P., & Tenhaken, R. (2006). Isolation of a novel ABC-transporter gene from soybean induced by salicylic acid. *Journal of Experimental Botany*, *57*, 2193–2201.

Ellis, R. J. (1979). The most abundant protein in the world. *Trends in Biochemical Sciences*, *4*, 241–244.

Emanuelsson, O., Brunak, S., von Heijne, G., & Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. *Nature Protocols*, *2*, 953–971.

Escamilla-Trevino, L. L., Chen, W., Card, M. L., Shih, M. C., & Poulton, J. E. (2006). *Arabidopsis thaliana b*-Glucosidases BGLU45 and BGLU46 hydrolyse monolignol glucosides. *Phytochemistry*, *67*, 1651–1660.

Eulgem, T. (2006). Dissecting the WRKY web of plant defense regulators. *PLOS Pathogens*, *2*, 126-134.

Everett, K. R., Pushparajah, I. P. S., Janssen, B. J., & Luo, Z. (2010). Advantages and disadvantages of microarrays to study microbial population dynamics – a minireview. *Society*, *6*, 1–6.

Fang, X., Chen, W., Xin, Y., Zhang, H., Yan, C., Yu, H., Ruan, S. (2012). Proteomic analysis of strawberry leaves infected with *Colletotrichum fragariae*. *Journal of Proteomics*, *75*, 4074–4090.

Felton, G. W., Bi, J. L., Mathews, M. C., Murphy, J. B., Korth, K., Wesley, S. V., Dixon, R. A. (1999). Cross-talk between the signal pathways for pathogen-induced systemic acquired resistance and grazing-induced insect resistance. *Novartis Foundation Symposium*, *223*, 166–171.

Felton, G. W., Korth, K. L., Bi, J. L., Wesley, S. V., Huhman, D. V., Mathews, M. C., Dixon, R. A. (1999). Inverse relationship between systemic resistance of plants to microorganisms and to insect herbivory. *Current Biology*, *9*, 317–320.

Fenn, J., Mann, M., Meng, C., Wong, S., & Whitehouse, C. (1989). Electrospray ionization for mass spectrometry of large biomolecules. *Science*, (246), 635–660.

Finehout, E. J., Franck, Z., Choe, L. H., Relkin, N., & Lee, K. H. (2007). Cerebrospinal fluid proteomic biomarkers for Alzheimer's disease. *Annals of Neurology*, *61*, 120–129.

Frazee, A. C., Johns, T., Street, N. W., Hansen, K. D., Irizarry, R. A., & Leek, J. T. (2014). Differential expression analysis of RNA-seq data at single-base resolution, *Biostatistics*, *15*, 413–426.

Fristensky, B., Horovitz, D., & Hadwiger, L. (1988). cDNA sequences for pea disease resistance response genes. *Plant Molecular Biology*, *11*, 713–715.

Fuller, M., Hamed, F., Wisniewski, M., & Glen, D. (2003). Protection of plants from frost using hydrophobic particle film and acrylic polymer. *Annals of Applied Biology*, *143*, 93–97.

Gao, F., Zhou, Y. J., Zhu, W. P., Li, X. F., Fan, L. M., & Zhang, G. F. (2009). Proteomic analysis of cold stress-responsive proteins in *Thellungiella* rosette leaves. *Planta*, *230*, 1033–1046.

Garofalo, C. G., Garavaglia, B. S., Dunger, G., Gottig, N., Orellano, E. G., & Ottado, J. (2009). Expression analysis of small heat shock proteins during compatible and incompatible plant-pathogen interactions, *Advanced Studies in Biology*, *1*, 197–205.

Gau, A. E., Koutb, M., Piotrowski, M., & Kloppstech, K. (2004). Accumulation of pathogenesis-related proteins in the apoplast of a susceptible cultivar of apple (*Malus* x *domestica* cv . Elstar) after infection by *Venturia inaequalis* and constitutive expression of PR genes in the resistant cultivar Remo. *European Journal of Plant Pathology*, *110*, 703–711.

Gelhaye, E., Rouhier, N., Navrot, N., & Jacquot, J. P. (2005). The plant thioredoxin system. *Cellular and Molecular Life Sciences*, *62*, 24–35.

Gene Ontology Consortium. (2000). Gene ontology: tool for the unification of biology. *Nature Genetics*, *25*, 25–29.

Gessler, C., Patocchi, A., Sansavini, S., Tartarini, S., & Gianfranceschi, L. (2006). *Venturia inaequalis* resistance in apple. *Critical Reviews in Plant Sciences, 25*, 473–503.

Gilmore, J. M., & Washburn, M. P. (2010). Advances in shotgun proteomics and the analysis of membrane proteomics. *Journal of Proteomics, 73*, 2078–2091.

Gladieux, P., Caffier, V., Devaux, M., Le Cam, B., & Cam, B. Le. (2010). Host-specific differentiation among populations of *Venturia inaequalis* causing scab on apple, pyracantha and loquat. *Fungal Genetics Biology*, *47*, 511–521.

Gladieux, P., Zhang, X-G., Afoufa-Bastien, D., Sanhueza, R-M. V., Sbaghi, M., Le Cam, B., Cam, B. Le. (2008). On the origin and spread of the scab disease of apple: out of central Asia. *PLOS One*, *3*, 1–14.

Gomez, C., Brun, L., Chauffour, D., & Valle, D. D. Le. (2007). Effect of leaf litter management on scab development in an organic apple orchard. *Crop Protection*, *118*, 249–255.

Gordo, S. M. C., Pinheiro, D. G., Moreira, E. C. O., Rodrigues, S. M., Poltronieri, M. C., de Lemos, O. F., Darnet, S. (2012). High-throughput sequencing of black pepper root transcriptome. *Biomed Central Plant Biology*, *12*, 168-177.

Granier, F. (1988). Extraction of plant proteins for two-dimensional electrophoresis. *Electrophoresis, 9*, 712–718.

Grant, M. R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Dangl, J. L. (1995). Structure of the *Arabidopsis* RPM1 gene enabling dual specificity disease resistance. *Science*, *269*, 843–846.

Griffith, M., Griffith, O. L., Mwenifumbo, J., Goya, R., Morrissy, A. S., Morin, R. D., Marra, M. A. (2010). Alternative expression analysis by RNA sequencing. *Nature Methods*, *7*, 843–847.

Guerin, F., Franck, P., Loiseau, A., Devaux, M., Le Cam, B., & Note, P. (2004). Isolation of 21 new polymorphic microsatellite loci in the phytopathogenic fungus *Venturia inaequalis*. *Molecular Ecology Notes*, *4*, 268–270.

Gusberti, M., Gessler, C., & Broggini, G. A. L. (2013). RNA-seq analysis reveals candidate genes for ontogenic resistance in *Malus-Venturia* pathosystem. *PLOS One*, *8*, 1-14.

Hajheidari, M., Abdollahian-noghabi, M., Askari, H., Heidari, M., Sadeghian, S. Y., Ober, E. S., & Salekdeh, G. H. (2005). Proteome analysis of sugar beet leaves under drought stress. *Proteomics*, *5*, 950–960.

Han, J. (1990). Use of antitranspirant epidermal coatings for plant protection in China. *Plant Disease*, *74*, 263–266.

Han, Q., Kang, G., & Guo, T. (2013). Plant Physiology and Biochemistry Proteomic analysis of spring freeze-stress responsive proteins in leaves of bread wheat (*Triticum aestivum* L .). *Plant Physiology and Biochemistry*, *63*, 236–244.

Hancock, J. ., Desikan, R., Clarke, A., Hurst, A., & Neill, S. (2002). Cell signalling following plant/pathogen interactions involves the generation of reactive oxygen and reactive nitrogen species. *Plant Physiology and Biochemistry*, *40*, 611–617.

Hancock, J. F. (2004). Plant Evolution and the Origin of Crop Species. *Wallingdorf: CABI Publishing.* pp 313-315.

Hardcastle, T. J., & Kelly, K. A. (2010). baySeq: Empirical Bayesian analysis of patterns of differential expression in count data. *Biomed Central Bioinformatics*, *11*, 422–436.

Harris, M. , Clark, J., Ireland, A., Lomax, J., & Ashburner, M. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Research*, *32*, 258–261.

Harris, S. A., P, J., Juniper, B. E., Harris, A., & Robinson, J. P. (2002). Genetic clues of the origin of the apple. *Trends in Genetics*, *18*, 426–430.

He, Z. H., He, D. Z., & Kohorn, B. D. (1998). Requirement for the induced expression of a cell wall associated receptor kinase for survival during the pathogen response. *The Plant Journal*, *14*, 55–63.

Heldt, H. . (1997). *Plant biochemistry and molecular biology*. Oxford: Oxford University Press. pp 26-33.

Holb, I. J., Heijne, B., Withagen, J. C. M., Gall, J., & Jeger, M. J. (2005). Analysis of summer epidemic progress of apple scab in different apple production systems in The Netherlands and Hungary. *Phytopathology*, *95*, 1001–1020.

Holb, I. J., Heijne, B., & Jeger, M. J. (2005). The widespread occurrence of overwintered conidial inoculum of *Venturia inaequalis* on shoots and buds in organic and integrated apple orchards across the Netherlands. *European Journal of Plant Pathology*, *111*, 157–168.

Holb, I. J., Heijne, B., & Jeger, M. J. (2006). Effects of integrated control measures on earthworms, leaf litter and *Venturia inaequalis* infection in two European apple orchards. *Agriculture Ecosystems and Environment*, *114*, 287–295.

Horton, P., Park, K-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J., & Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, *35*, W585–W587.

Hough, L., Shay, J., & Dayton, D. (1953). Apple scab resistance from *Malus floribunda* Sieb. *Proceedings of the American Society of Horticultural Sciences*, *62*, 341–347.

Huang, X. L., Liu, L. X., Chen, J., & Zhai, Y. H. (2009). Comparative proteomic analysis of the response in resistant and susceptible maize inbred lines to infection by *Curvularia lunata*. *Progress in Natural Science*, *19*, 845–850.

Ifuku, K., Ido, K., & Sato, F. (2011). Molecular functions of PsbP and PsbQ proteins in the photosystem II supercomplex. *Journal of Photochemistry and Photobiology*, *104*, 158–64.

Jain, R., Katavic, V., Agrawal, G. K., Guzov, V. M., & Thelen, J. J. (2008). Purification and proteomic characterization of plastids from *Brassica napus* developing embryos. *Proteomics*, *8*, 3397–3405.

Janick, J., Cummins, J. N., Brown, S. K., & Hemmat, M. (1996). Apples. *In: Janick J and Moore JN (eds) Fruit Breed, Volume 1: Tree and Tropical Fruits. New York: John Wiley & Sons*, pp 1–77.

Jha, G., Thakur, K., & Thakur, P. (2010). The *Venturia* apple pathosystem: pathogenicity mechanisms and plant defense responses. *Journal of Biomedicine and Biotechnology*, *2009*, 1–10.

Kalde, M., Nuhse, T. S., Findlay, K., & Peck, S. C. (2007). The syntaxin SYP132 contributes to plant resistance against bacteria and secretion of pathogenesis-related

protein 1. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 11850–11855.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M.,Yamanishi, Y. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Research*, *36*, 480–484.

Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopaedia of Genes and Genomes. *Nucleic Acids Research*, *28*, 27–30.

Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K., Itoh, M., Kawashima, S., Hirakawa, M. (2006). From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research*, *34*, 354–357.

Kang, G., Li, G., Zheng, B., Han, Q., Wang, C., Zhu, Y., & Guo, T. (2012). Proteomic analysis on salicylic acid-induced salt tolerance in common wheat seedlings (*Triticum aestivum L* .). *Proteins and Proteomics*, *1824*, 1324–1333.

Kanzaki, H., Saitoh, H., Ito, A., Fujisawa, S., Kamoun, S., Katou, S., Terauchi, R. (2003). Cytosolic HSP90 and HSP70 are essential components of INF1-mediated hypersensitive response and non-host resistance to *Pseudomonas cichorii* in *Nicotiana benthamiana*. *Molecular Plant Pathology*, *4*, 385–391.

Karas, M., & Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10000 Daltons. *Analytical Chemistry*, (60), 2299–2301.

Katam, R., Basha, S. M., Suravajhala, P., & Pechan, T. (2010). Analysis of Peanut Leaf Proteome research articles. *Journal of Proteome Research*, 2236–2254.

Kawahara, Y., Oono, Y., Kanamori, H., Matsumoto, T., Itoh, T., & Minami, E. (2012). Simultaneous RNA-seq analysis of a mixed transcriptome of rice and blast fungus interaction. *PLOS One, 7*, 1-15.

Kawano, T. (2003). Roles of the reactive oxygen species-generating peroxidase reactions in plant defense and growth induction. *Plant Cell Reports, 21*, 829–837.

Kayim, M., Ceccardi, T. L., Berretta, M. J., Barthe, G. A., & K.S., D. (2004). Introduction of a citrus blight-associated gene into Carrizo citrange [*Citrus sinensis* (L.) Osbc. x *Poncirus trifoliata* (L.) Raf.] by *Agrobacterium*-mediated transformation. *Plant Cell Reports, 23*, 377–385.

Kedzierska, S. (2006). Structure, function and mechanisms of action of ATPases from the AAA superfamily of proteins. *Postepy Biochemii, 52*, 330–338.

Kent, W. (2002). BLAT - the BLAST-like alignment tool. *Genome Research*, *12*, 656–664.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, *14*, 36-47.

Kim, S. T., Cho, K. S., Yu, S., Kim, S. G., Hong, J. C., Han, C. D., Kang, K. Y. (2003). Proteomic analysis of differentially expressed proteins induced by rice blast fungus and elicitor in suspension-cultured rice cells. *Proteomics, 3*, 2368–2378.

Kim, S. T. G., Yu, S., Kang, S. Y., Kim, H. J., Hwang, D. H., & Kang, K. Y. (2004). Identification and molecular characterizaton of rice blast fungus *M-grisea* induced proteins through proteomic analysis. *Phytopathology*, *94*, 52–65.

Kim, Y. K., & Lee, M. Y. (2009). Proteomic analysis of differentially expressed proteins of rice in response to cadmium. *Journal of the Korean Society for Applied Biological Chemistry*, *52*, 428–436.

Knoth, C., Ringler, J., Dangl, J. L., & Eulgem, T. (2007). *Arabidopsis* WRKY70 is required for full RPP4-mediated disease resistance and basal defense against *Hyaloperonospora parasitica*. *Molecular Plant Microbe Interaction*, *20*, 120–128.

Kohorn, B. D., Johansen, S., Shishido, A., Todorova, T., Martinez, R., Defeo, E., & P, O. (2009). Pectin activation of MAP kinase and gene expression is *WAK2* dependent. *Plant Journal*, *60*, 974–982.

Koller, W. (1994). Chemical control of apple scab-*status quo* and future. *Norwegian Journal of Agricultural Sciences. Supplement.*, *17*, 149–170.

Koller, W., Parker, D. M., & Becker, C. M. (1991). Role of cutinase in the penetration of apple leaves by *Venturia inaequalis*. *Phytopathology*, *81*, 1375–1379.

Konishi, H., Ishiguro, K., & Komatsu, S. (2001). A proteomics approach towards understanding blast fungus infection of rice grown under different levels of nitrogen fertilization. *Proteomics*, *1*, 1162–1171.

Koorneef, A., & Pieterse, C. M. (2008). Cross talk in defense signalling. *Plant Physiology*, *146*, 839–844.

Korban, S. S., & Skirvin, R. M. (1984). Nomeclature of the cultivated apple. *Hortscience*, *19*, 177–180.

Kristoffersen, P., Brzobohaty, B., Hohfeld, I., Bako, L., Melkonian, M., & Palme, K. (2000). Developmental regulation of the maize Zm-g60. 1 gene encoding a beta-glucosidase located to plastids. *Planta*, *210*, 407–415.

Kucheryava, N., Bowen, J. K., Sutherland, P. W., Conolly, J. J., Mesarich, C. H., Rikkerink, E. H., Templeton, M. D. (2008). Two novel *Venturia inaequalis* genes induced upon morphogenetic differentiation during infection and *in vitro* growth on cellophane. *Fungal Genetics Biology*, *45*, 1329–1339.

Kumar, A., Agarwal, S., Heyman, J. A., Matson, S., Heidtman, M., Piccirillo, S., Snyder, M. (2002). Sub cellular localization of the yeast proteome. *Genes and Development*, *16*, 707–719.

Kuniak, E., & Sklodowska, M. (2001). Ascorbate, glutathione and related enzymes in chloroplasts of tomato leaves infected by *Botrytis cinerea*. *Plant Science*, *160*, 723–731.

Kuzniak, E. (2010). The ascorbate-glutathione cycle and related redox signals in plant-pathogen interactions. *Ascorbate-Glutathione Pathway and Stress Tolerance in Plants*, *Dordrecht: Springer*, pp 115-123.

Lamb, C., & Dixon, R. (1997). The oxidative burst in plant disease resistance. *Annual Review of Plant Physiology and Plant Molecular Biology*, *48*, 251–275.

Langmead, B., Trapnell, C., Mihai, P., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, *10*, 25-34.

Lee, J., Feng, J., Campbell, K. B., Scheffler, B. E., Garrett, W. M., Thibivilliers, S., Cooper, B. (2009). Quantitative Proteomic Analysis of Bean Plants Infected by a Virulent and Avirulent Obligate Rust Fungus. *Molecular & Cellular Proteomics*, *8*, 19–31.

Lévesque, C. A., Brouwer, H., Cano, L., Hamilton, J. P., Holt, C., Huitema, E., Buell, C. R. (2010). Genome sequence of the necrotrophic plant pathogen *Pythium ultimum* reveals original pathogenicity mechanisms and effector repertoire. *Genome Biology*, *11*, 73-84.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics, 25*, 1754–1760.

Li, J., Brader, G., & Palva, E. T. (2004). The WRKY70 transcription factor: A node of convergence for jasmonate-mediated and salicylate-mediated signals in plant defense. *Plant Cell*, *16*, 319–331.

Li, R., Yu, C., Li, Y., Lam, T-W., Yiu, S-M., Kristiansen, K., & Wang, J. (2009). SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, *25*, 1966–1967.

Li, X., Bai, T., Li, Y., Ruan, X., & Li, H. (2013). Proteomic analysis of *Fusarium oxysporum f. sp. cubense* tropical race 4-inoculated response to *Fusarium* wilts in the banana root cells. *Proteome Science*, *11*, 41.

Liang, Y., Srivastava, S., Rahman, M. H., Strelkov, S. E., & Kav, N. N. V. (2008). Proteome changes in leaves of *Brassica napus* L. as a result of *Sclerotinia sclerotiorum* challenge. *Journal of Agricultural and Food Chemistry*, *56*, 1963–1976.

Lin, Y. Z., Chen, H. Y., Kao, R., Chang, S. P., Chang, S. J., & Lai, E. M. (2008). Proteomic analysis of rice defense response induced by probenazole. *Phytochemistry*, *69*, 715–728.

Liu, H-Y., Dai, J-R., Feng, D-R., Liu, B., Wang, H-B., & Wang, J-F. (2010). Characterization of a novel plantain *Asr* gene, *MpAsr*, that is regulated in response to infection of *Fusarium oxysporum f. sp. cubense* and abiotic stresses. *Journal of Integrative Plant Biology*, *52*, 315–23.

Liu, Z. Y., & Halterman, D. (2009). Analysis of proteins differentially accumulated during potato late blight resistance mediated by the RB resistance gene. *Physiological and Molecular Plant Pathology*, *74*, 151–160.

Lolis, E., Alber, R., Davenport, R. C., Rose, D., Hartman, F. C., & Petsko, G. A. (1990). Structure of Yeast Triosephosphate Isomerase at 1.9-Å Resolution. *American Chemical Society*, *29*, 6609–6618.

Luche, S., Lelong, C., Diemer, H., Van Dorsselaer, A., & Rabilloud, T. (2007). Ultrafast coelectrophoretic fluorescent staining of proteins with carbocyanines. *Proteomics*, *7*, 3234–3244.

Machardy, W. E. (1996). The disease. *In: MacHardy WE (ed) Apple Scab: Biology, Epidemiology and Managament. APS Press, St. Paul, Minnesota,* pp *1-28.*

MacHardy, W. E., Gadoury, D. M., & Gessler, C. (2001). Parasitic and Biological Fitness of *Venturia inaequalis*: Relationship to Disease Management Strategies. *Plant Disease*, *85*, 1036–1051.

Mahmood, T., Jan, A., Kakishima, M., & Komatsu, S. (2006). Proteomic analysis of bacterial-blight defense-responsive proteins in rice leaf blades. *Proteomics*, *6*, 6053–6065.

Maimbo, M., Ohnishi, K., Hikichi, Y., Yoshioka, H., & Kiba, A. (2007). Induction of a small heat shock protein and its functional roles in *Nicotiana* plants in the defense response against *Ralstonia solanacearum*. *Plant Physiology, 145*, 1588–1599.

Malnoy, M., Xu, M., Borejsza-Wysocka, E., Korban, S. S., & Aldwinckle, H. S. (2008). Two receptor-like genes, *Vfa1* and *Vfa2*, confer resistance to the fungal pathogen *Venturia inaequalis* inciting apple scab disease. *Molecular Plant Microbe Interaction*, *21*, 448–458.

Mann, K., & Mann, M. (2011). In-depth analysis of the chicken egg white proteome using an LTQ Orbitrap Velos. *Proteome Science*, *9*, 7-15.

Maor, R., Puyesky, M., Horwitz, B., & Sharon, P. (1998). Use of green fluorescent protein (GFP) for studying development and fungal-plant interaction in *Cochliobolus heterostrophus*. *Mycological Research*, *102*, 491–496.

Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., & Gilad, Y. (2008). RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, *18*, 1509–1517.

Marmagne, A., Ferro, M., Meinnel, T., Bruley, C., Kuhn, L., Garin, J., Ephritikhine, G. (2007). A high content in lipid- modified peripheral proteins and integral receptor kinases features in the *Arabidopsis* plasma membrane proteome. *Molecular and Cellular Proteomics*, *6*, 1980–1996.

Marra, R., Li, H., Barbetti, M. J., Sivasithamparam, K., Vinale, F., Cavallo, P., & Lorito, M. (2010). Proteomic analysis of the interaction between *Brassica napus* cv. Surpass 400 and virulent or avirulent isolates of *Leptosphaeria maculans*. *Journal of Plant Pathology*, *92*, 89–101.

Martin, J. A, & Wang, Z. (2011). Next-generation transcriptome assembly. *Nature Reviews Genetics*, *12*, 671–682.

Martínez-Gómez, P., Crisosto, C. H., Bonghi, C., & Rubio, M. (2011). New approaches to *Prunus* transcriptome analysis. *Genetica*, *139*, 755–769.

Mcgettigan, P. A. (2012). Transcriptomics in the RNA-seq era. *Current Opinion in Chemical Biology*, *17*, 1-8.

Mckay, S. J., Paul, S., Bradeen, J. M., & Luby, J. J. (2011). Prediction of genotypic values for apple fruit texture traits in a breeding population derived from "Honeycrisp ". *Distribution*, *136*, 408–414.

Menhaj, A. R., Mishra, S. K., Bezhani, S., & Kloppstech, K. (1999). Posttranscriptional control in the expression of the genes coding for high-light-regulated HL#2 proteins. *Planta*, *209*, 406–13.

Miernyk, J. A. (1997). The 70 kDa stress-related proteins as molecular chaperones. *Trends in Plant Science*, *2*, 180–187.

Mittler, R., Vanderauwera, S., Gollery, M., & Breusegem, F. (2004). Reactive oxygen gene networks of plants. *Trends in Plant Science*, *9*, 490–498.

Morant, A. V., Jørgensen, K., Jørgensen, C., Paquette, S. M., Sánchez-Pérez, R., Møller, B. L., & Bak, S. (2008). beta-Glucosidases as detonators of plant chemical defense. *Phytochemistry*, *69*, 1795–1813.

Mortazavi, A., Williams, B. A., Mccue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods*, *5*, 621–628.

Murphy, J. P., Kong, F. M., Pinto, D. M., & Wang-Pruski, G. (2010). Relative quantitative proteomic analysis reveals wound response proteins correlated with after-cooking darkening. *Proteomics*, *10*, 4258–4269.

Nagalakshmi, U., Waern, K., & Snyder, M. (2010). RNA-Seq : A Method for Comprehensive Transcriptome Analysis. *Current Protocols in Molecular Biology*, *5*, 1–13.

Nanjo, Y., Skultety, L., Ashraf, Y., & Komatsu, S. (2010). Comparative Proteomic Analysis of Early-Stage Soybean Seedlings Responses to Flooding by Using Gel and Gel-Free Techniques. *Journal of Proteome Research*, *9*, 3989–4002.

Nelson, T., Tausta, S. L., Gandotra, N., & Liu, T. (2006). Laser microdissection of plant tissue: What you see is what you get. *Annual Review of Plant Biology*, *57*, 181–201.

Neuhoff, V., Arold, N., Taube, D., & Ehrhardt, W. (1988). Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis*, *9*, 255–62.

Ngara, R., Ndimba, R., Borch-Jensen, J., Jensen, N., & Ndimba, B. (2012). Identification and profiling of salinity stress-responsive proteins in *Sorghum bicolor* seedlings. *Journal of Proteomics*, *75*, 4139–4150.

Nicholson, R. L., Kuc, J., & Williams, B. (1972). Histochemical demonstration of transitory esterase activity in *Venturia inaequalis*. *Phytopathology*, *62*, 1242–1247.

Niklas, J., & Kennel, W. (1981). The role of the earthworm, *Lumbricus terrestris* (L) in removing sources of phytopathogenic fungi in orchards. *Gartenbauwissenschaft*, *46*, 138–142.

Nürberger, T., & Kernmerling, B. (2006). Receptor protein-kinases-pattern recognition receptors in plant immunity. *Trends in Plant Science*, *11*, 519–522.

Nusbaum, C. J., & Keitt, G. W. (1938). A cytological study of host-parasite relations of *Venturia inaequalis* on apple leaves. *Journal of Agricultural Research*, *56*, 595–618.

O'Farrel, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, *250*, 4007–4021.

Oshlack, A., Robinson, M. D., & Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biology*, *11*, 220.

Oshlack, A., & Wakefield, M. J. (2009). Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, *4*, 14-19.

Osswald, W., Niehuss, M., Huber, W., & Elstner, E. . (1984). Support of non-host resistance by artificial leaf coating. *Journal of Plant Diseases and Protection*, *91*, 337–344.

Parisi, L., Fouillet, V., Schouten, H. J., Groenwold, R., Laurens, F., Didelot, F., Thissen, J., and Tsipouridis, C. (2004). Variability of the pathogenicity of *Venturia inaequalis* in Europe. *Acta Horticulturae*, *663*, 107–114.

Parisi, L., Lespinasse, Y., Guillaumes, J., & Kilger, J. (1993). A new race of *Venturia inaequalis* virulent to apples with resistance due to the Vf gene. *Phytopathology*, *83*, 533–537.

Parker, J. E., Holub, E. B., Frost, L. N., Falk, A., Gunn, N. D., & Daniels, M. J. (1996). Characterization of *eds1*, a mutation in *Arabidopsis* suppressing resistance to *Peronospora parasitica* specified by several different RPP genes. *Plant Cell*, *11*, 2033–2046.

Patocchi, A., Gianfranceschi, L., & Gessler, C. (1999). Towards the map-based cloning of *Vf*: Fine and physical mapping of the *Vf* region. *Theoretical and Applied Genetics*, *99*, 1012–1017.

Peng, Z. Y., Wang, M. C., Li, F., Lv, H. J., Li, C. L., & Xia, G. M. (2009). A proteomic study of the response to salinity and drought stress in an introgression strain of bread wheat. *Molecular and Cellular Proteomics*, *8*, 2676–2686.

Percival, G. C., & Boyle, S. (2009). Evaluation of film forming polymers to control apple scab (*Venturia inaequalis* (Cooke) Wint.) under laboratory and field conditions. *Crop Protection*, *28*, 30–35.

Pickelsimer, C. (2013). South Africa - Republic of Fresh Deciduous Fruit Annual South African Deciduous Exports Increase on a Weak Rand. *Fresh*, 1-17.

Pieterse, C. M., & Van Loon, L. C. (2004). NPR1: the spider in the web of induced resistance signaling pathways. *Current Opinion in Plant Biology*, *7*, 456–464.

Pnueli, L., Liang, H., Rozenberg, M., & Mittler, R. (2003). Growth suppression, altered stomatal responses, and augmented induction of ehat shock proteins in cytosolic ascorbate peroxidase (Apx1)-deficient *Arabidopsis* plants. *Plant Journal*, *34*, 187–203.

Popelkova, H., & Yocum, C. F. (2011). PsbO, the manganese-stabilizing protein: analysis of the structure-function relations that provide insights into its role in photosystem II. *Journal of Photochemistry and Photobiology. 104*, 179–90.

Pritsch, C., Muehlbauer, G. J., Bushnell, W. R., Somers, D. A, & Vance, C. P. (2000). Fungal development and induction of defense response genes during early infection

of wheat spikes by *Fusarium graminearum*. *Molecular Plant-Microbe Interactions*, *13*, 159–69.

Proietti, S., Bertini, L., Timperio, A. M., Zolla, L., Caporale, C., & Caruso, C. (2013). Crosstalk between salicylic acid and jasmonate in *Arabidopsis* investigated by an integrated proteomic and transcriptomic approach. *Molecular bioSystems*, *9*, 1169–1187.

Quirino, B. F., Candido, E. S., Campos, P. F., Franco, O. L., Krüger, R. H., & Kruger, R. H. (2010). Proteomic approaches to study plant-pathogen interactions. *Phytochemistry*, *71*, 351–362.

R Development Core Team. (2011). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing, Vienna, Austria*.

Rabilloud, T., & Lelong, C. (2011). Two-dimensional gel electrophoresis in proteomics: A tutorial. *Journal of Proteomics*, *74*, 1829–1841.

Radauer, C., Lackner, P., & Breiteneder, H. (2008). The Bet v 1 fold: an ancient, versatile scaffold for binding of large, hydrophobic ligands. *Biomed Central Evolutionary Biology*, *8*, 286-294.

Robinson, M., McCarthy, D. J., & Smyth, G. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*, 139–140.

Robinson, M., & Smyth, G. (2007). Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics*, *23*, 2881–2887.

Robinson, M., Mccarthy, D., Chen, Y., & Smyth, G. K. (2013). edgeR: differential expression analysis of digital gene expression data User's Guide, (March).

Rushton, P. J., Somssich, I. E., Ringler, P., & Shen, Q. J. (2010). WRKY transcription factors. *Trends in Plant Science*, *15*, 247–258.

Rushton, P. J., Torres, J., Parniske, M., Wernert, P., Hahlbrock, K., & Somssich, I. E. (1996). Interaction of elicitor-induced DNA-binding proteins with elicitor response elements in the promoters of parsley PR1 genes. *Embo Journal*, *15*, 5690–5700.

Ryu, H-S., Song, M-Y., Kim, C-Y., Muho, H., Ryoo, N., Lee, S-K., Jeon, J-S. (2009). Proteomic analysis of rice mutants susceptible to *Magnaporthe oryzae*. *Plant Biotechnology Reports, 3*, 167–174.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*, 5463–5467.

Savory, E. A., Adhikari, B. N., Hamilton, J. P., Vaillancourt, B., Buell, C. R., & Day, B. (2012). mRNA-Seq analysis of the *Pseudoperonospora cubensis* transcriptome during cucumber (*Cucumis sativus* L.) infection. *PLOS One, 7*, 1-12.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantative monitoring of gene expression patterns with a complementary DNA microarray. *Science of the Total Environment*, *270*, 467–470.

Schenk, S., Schoenhals, G. J., Souza, G. D., & Mann, M. (2008). A high confidence, manually validated human blood plasma protein reference set. *Biomed Central Medical Genomics*, 5, *1:41*.

Schouten, H. J., Krens, F. ., & Jacobsen, E. (2006). *Cis*-genic plants are similar to traditionally bred plants. *Embo Reports*, *7*, 750–53.

Schumaker, C. F. A., Steiner, U., Dehne, H-W., & Oerke, E-C. (2008). Localized adhesion of nongerminated *Venturia inaequalis* conidia to leaves and artificial surfaces. *Phytopathology*, *98*, 760–768.

Seki, M., Narusaka, M., Kamiya, A., Ishida, J., Satou, M., Sakurai, T., Shinozaki, K. (2002). Functional Annotation of a Full-Length *Arabidopsis* cDNA Collection. *Science, 296,* 141–145.

Sharma, N., Rahman, M. H., Strelkov, S., Thiagarajah, M., Bansal, V. K., & Kav, N. N. V. (2007). Proteome-level changes in two *Brassica napus* lines exhibiting differential responses to the fungal pathogen *Alternaria brassicae*. *Plant Science*, *172*, 95–110.

Shi, C-Y., Yang, H., Wei, C-L., Yu, O., Zhang, Z-Z., Jiang, C-J., Wan, X-C. (2011). Deep sequencing of the *Camellia sinensis* transcriptome revealed candidate genes for major metabolic pathways of tea-specific compounds. *Biomed Central Genomics*, *12*, 131-142.

Shin, K., Hena, A., Kamal, M., Cho, K., Choi, J., Lee, Y. W., Heo, H. (2011). Defense proteins are induced in wheat spikes exposed to *Fusarium graminearum*, *4*, 270–277.

Siphugu, L. (2011). South Africa - Republic of Fresh Deciduous Fruit Semi-annual. *Agricultural Marketing*. *Fresh,* 1-18.

Smereka, K. J., Machardy, W. E., & Kausch, A. P. (1987). Cellular in *Venturia inaequalis* ascospores during germination and penetration of apple leaves. *Canadian Journal of Botany*, *65*, 2549–2561.

Smyth, G. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Statistical Application in Genetics and Molecular Biology*, *3*, 1-3.

Sobhanian, H., Razavizadeh, R., Nanjo, Y., Ehsanpour, A. A., Jazii, F. R., Motamed, N., & Komatsu, S. (2010). Proteome analysis of soybean leaves, hypocotyls and roots under salt stress. *Proteome Science*, *8*, 1-15.

Socquet-Juglard, D., Kamber, T., Pothier, J. F., Christen, D., Gessler, C., Duffy, B., & Andrea, P. (2013). Comparative RNA-Seq Analysis of Early-Infected Peach Leaves by the Invasive Phytopathogen *Xanthomonas arboricola pv. pruni*. *PLOS One*, *8*, 1–12.

Soriano, J., Joshi, S., Van Kaauwen, M., Noordijk, Y., Groenwold, R., Henken, B., Schouten, H. J. (2009). Identification and mapping of the novel apple scab resistance gene Vd3. *Tree Genetics and Genomes*, *5*, 475–482.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology*, *98*, 503–517.

Srivastava, S., & Chen, L. A. (2010). A two-parameter generalized Poisson model to improve the analysis of RNA-seq data. *Nucleic Acids Research*, *38*, 1-15.

Stukkens, Y., Bultreys, A., Grec, S., Trombik, T., Vanham, D., Boutry, M., & M. (2005). NpPDR1, a pleiotropic drug resistance-type ATP-binding cassette transporter from *Nicotiana plumbaginifolia*, plays a major role in plant pathogen defense. *Plant Physiology*, *139*, 341–352.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., & Ebert, B. L. (2005). Gene set enrichment analysis : A knowledge-based approach for interpreting genome-wide. *Proceedings of the National Academy of Sciences*, *102*, 15545-15550.

Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Yaspo, M. L. (2008). A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science*, *321*, 956–960.

Sutherland, F., & Walters, D. . (2001). *In vitro* effects of film-forming polymers on the growth and morphology of *Pyrenophora avenae* and *Pyricularia oryzae*. *Journal of Phytopathology*, *149*, 621–624.

Sutherland, F., & Walters, D. (2002). Effect of film forming polymers on infection of barley with the powdery mildew fungus, *Blumeria graminis f. sp. hordei*. *European Journal of Plant Pathology*, *108*, 385–389.

Szankowski, I., Waidmann, S., Degenhardt, J., Patocchi, A., Paris, R., Silfverberg-Dilworth, E., Gessler, C. (2009). Highly scab-resistant transgenic apple lines achieved by introgression of *HcrVf2* controlled by different native promoter lengths. *Tree Genetics and Genomes*, *5*, 349–358.

Tang, W., Coughlan, S., Crane, E., Beatty, M., & Duvick, J. (2006). The application of laser microdissection to in planta gene expression profiling of the maize anthracnose stalk rot fungus *Colletotrichum graminicola*. *Molecular Plant-Microbe Interactions*, *19*, 1240–1250.

Tang, W.-H., Zhang, Y., & Duvick, J. (2012). The application of laser microdissection to profiling fungal pathogen gene expression *in planta*. *Methods in Molecular Biology*, *835*, 219–236.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Research, 17,* 6463–6471.

Thakur, K., Chawla, V., Bhatti, S., Swarnkar, M. K., Kaur, J., Shankar, R., & Jha, G. (2013). *De novo* transcriptome sequencing and analysis for *venturia inaequalis*, the devastating apple scab pathogen. *PLOS One, 8,* 1-16.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics, 25,* 1105–11.

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology*, *28,* 511–515.

Turcatti, G., Romieu, A., Fedurco, M., & Tairi, A-P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. *Nucleic Acids Research, 36,* 25-29.

Valcu, C.-M., Junqueira, M., Shevchenko, A., & Schlink, K. (2009). Comparative proteomic analysis of responses to pathogen infection and wounding in *Fagus sylvatica*. *Journal of Proteome Research, 8,* 4077–4091.

Van der Biezen, E. A., & Jones, J. D. G. (1998). The NB-ARC domain: a novel signalling motif shared by plant resistance gene products and regulators of cell death in animals. *Current Biology*, *8,* 226–228.

Van Wijk, K. J. (2004). Plastid proteomics. *Plant Physiology and Biochemistry, 42,* 963–977.

Velasco, R., Zharkikh, A., Affourtit, J., Dhingra, A., Cestaro, A., Kalyanaraman, A., Viola, R. (2010). The genome of the domesticated apple (*Malus* × *domestica* Borkh.). *Nature Genetics*, *42*, 833–839.

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

Venu, R. C., Madhav, M. S., Sreerekha, M. V, Nobuta, K., Zhang, Y. A., Carswell, P., Wang, G. L. (2010). Deep and Comparative Transcriptome Analysis of Rice Plants Infested by the Beet Armyworm (*Spodoptera exigua*) and Water Weevil (*Lissorhoptrus oryzophilus*). *Rice*, *3*, 22–35.

Vincent, D., Rancourt, B., & Carisse, O. (2004). Apple leaf shredding as a non-chemical tool to managed apple scab and spotted tentiform leafminer. *Agriculture Ecosystems & Environment*, *104*, 595–604.

Voegele, R. T., & Mendgen, K. (2003). Rust haustoria: Nutrient uptake and beyond. *New Phytologist*, *159*, 93–100.

Wang, K., Singh, D., Zeng, Z., Coleman, S. J., Huang, Y., Savich, G. L., Liu, J. Z. (2010). MapSplice: Accurate mapping of RNA-seq reads for splice junction discovery. *Nucleic Acids Research*, *38*, 1-15.

Wang, L., Feng, Z., Wang, X., Wang, X., & Zhang, X. (2009). DEGseq : an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, *26*, 136-138.

Wang, W., Monica, S., Rita, V., Antonia, S., Elisabetta, S., Silvia, M., & Mauro, C. (2003). Protein extraction for two-dimensional electrophoresis from olive leaf, a

plant tissue containing high levels of interfering compounds. *Electrophoresis*, *24*, 2369–2375.

Wang, W., Qin, Z., Feng, Z., Wang, X., & Zhang, X. (2013). Identifying differentially spliced genes from two groups of RNA-seq samples. *Gene*, *518*, 164–70.

Wang, W., Vignani, R., Scali, M., & Cresti, M. (2006). A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis*, *27*, 2782–2786.

Wang, W., Vinocur, B., Shoseyov, O., & Altman, A. (2004). Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science*, *9*, 244–252.

Wang, Y., & Fristensky, B. (2001). Transgenic canola lines expressing pea defense gene DRR206 have resistance to aggressive blackleg isolates and to *Rhizoctonia solani*. *Molecular Breeding*, *8*, 263–271.

Wang, Y., Yang, L. M., Xu, H. B., Li, Q. F., Ma, Z. Q., & Chu, C. G. (2005). Differential proteomic analysis of proteins in wheat spikes induced by *Fusarium graminearum*. *Proteomics*, *5*, 4496–4503.

Wang, Y., Yuan, X., Hu, H., Liu, Y., Sun, W., Shan, Z., & Zhou, X. (2012). Proteomic analysis of differentially expressed proteins in resistant soybean leaves after *Phakopsora pachyrhizi* infection, *Journal of Phytopathology*, *560*, 554–560.

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*, 57–63.

Werck-Reichhart, D., & Feyereisen, R. (2000). Cytochrome P450: a success story. *Genome Biology 1: Reviews*, *1*, 1-9.

Westermeier, R. (2005). Electrophoresis in Practice. *Fourth Edition*, pp 23-31.

Wiermer, M., Feys, B. J., & Parker, J. E. (2005). Plant immunity: the *EDS1* regulatory node. *Current Opinion in Plant Biology*, *8*, 383–389.

Wisniewski, J. R., Zougman, A., Nagaraj, N., & Mann, M. (2009). Universal sample preparation method for proteome analysis, *6*, 359–361.

Wood, F. (2003). The Silk Road. *London: The Folio Society.* pp *270*-277.

Wu, J., Chen, Y., & Gu, W. (2010). Urinary proteomics as a novel tool for biomarker discovery in kidney diseases. *Journal of Zhejiang University. Science*, *11*, 227–237.

Wu, T. D., & Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, *26*, 873–881.

Xin, M., Wang, X., Peng, H., Yao, Y., Xie, C., Han, Y., Sun, Q. (2012). Transcriptome comparison of susceptible and resistant wheat in response to powdery mildew infection. *Genomics, Proteomics and Bioinformatics*, *10*, 94–106.

Xu, M., & Korban, S. S. (2002). A cluster of four receptor-like genes resides in the *Vf* locus that confers resistance to apple scab disease. *Genetics*, *162*, 1995–2006.

Xu, P., Wang, J., & Fincher, G. B. (1992). Evolution and differential expression of the (1-->3)-beta-glucan endohydrolase-encoding gene family in barley, *Hordeum vulgare*. *Gene*, *120*, 157–65.

Yadeta, K. A., Elmore, J. M., & Coaker, G. (2013). Advancements in the analysis of the *Arabidopsis* plasma membrane proteome. *Frontiers in Plant Science*, *4*, 1–10.

Yan, S. P., Zhang, Q. Y., Tang, Z. C., Su, W. A., & Sun, W. N. (2006). Comparative proteomic analysis provides new insight into chilling stress response in rice. *Molecular and Cell Proteomics*, *5*, 484–496.

Yang, F., Jensen, J. D., Svensson, B., Jørgensen, H. J. L., & Collinge, D. B. (2010). Analysis of early events in the interaction between *Fusarium graminearum* and the susceptible barley (*Hordeum vulgare*) cultivar Scarlett. *Biotechnology*, *10*, 1–8.

Yang, K., Rong, W., Qi, L., Li, J., Wei, X., & Zhang, Z. (2013). Isolation and characterization of a novel wheat cysteine-rich receptor-like kinase gene induced by *Rhizoctonia cerealis*. *Scientific Reports*, *3*, 1-15.

Yao, Y., Sun, H., Xu, F., Zhang, X., & Liu, S. (2011). Comparative proteome analysis of metabolic changes by low phosphorus stress in two *Brassica napus* genotypes. *Planta*, *233*, 523–537.

Yates, J. R., Ruse, C. I., & Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. *Annual Review of Biomedical Engineering*, *11*, 49–79.

Young, A. (2001). Hsp90: a specialized but essential protein-folding tool. *Journal of Cellular Physiology*, *154*, 267–273.

Young, M. D., Wakefield, M. J., Smyth, G. K., & Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology*, *11*, 14-19.

Yu, C. L., Yan, S. P., Wang, C. C., Hu, H. T., Sun, W. N., Yan, C. Q., Yang, L. (2008). Pathogenesis-related proteins in somatic hybrid rice induced by bacterial blight. *Phytochemistry*, *69*, 1989–1996.

Zdobnov, M., & Apweiler, R. (2001). InterProScan-an integration platform for the signature-recognition methods in InterPro. *Bioinformatics*, *17*, 847–848.

Zhang, H., Hu, Y. A., Wang, C. Y., & Ji, W. Q. (2011). Gene expression in wheat induced by inoculation with *Puccinia striiformis* West. *Plant Molecular Biology Reporter*, *29*, 458–465.

Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., & Yates, J. R. (2013). Protein Analysis by shotgun/bottom-up proteomics. *Chemical Reviews*, *113*, 2343–2394.

Zipfel, C., Kunze, G., Chinchilla, D., Caniard, A., Jones, J. D. G., Boller, T., & Felix, G. (2006). Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts *Agrobacterium*-mediated transformation. *Cell*, *125*, 749–760.

Zipfel, C., Robatzek, S., Navarro, L., Oakeley, E. J., Jones, J. D. G., Felix, G., & Boller, T. (2004). Bacterial disease resistance in *Arabidopsis* through flagellin perception. *Nature*, *428*, 764–767.

Appendix A

Table 4.3 KEGG pathways for up-regulated genes

Pathway	Seqs in Pathway	Enzyme	Ezyme ID	Seqs of Enzyme	Pathway Image
Cyanoamino acid metabolism	8	lyase	ec:4.1.2.10	7	map00460_20131216054258.png
Cyanoamino acid metabolism	8	beta-glucosidase	ec:3.2.1.117	1	map00460_20131216054258.png
Aminobenzoate degradation	14	hydratase	ec:4.2.1.17	4	map00627_20131216054304.png
Aminobenzoate degradation	14	phosphatase	ec:3.1.3.2	10	map00627_20131216054304.png
Flavone and flavonol biosynthesis	1	3'-monooxygenase	ec:1.14.13.21	1	map00944_20131216054308.png
Toluene degradation	4	dehydrogenase	ec:1.1.1.35	4	map00623_20131216054313.png
Nitrogen metabolism	7	synthase (NADH)	ec:1.4.1.14	1	map00910_20131216054321.png
Nitrogen metabolism	7	dehydrogenase [NAD(P)+]	ec:1.4.1.3	1	map00910_20131216054321.png
Nitrogen metabolism	7	dehydrogenase	ec:1.4.1.2	1	map00910_20131216054321.png
Nitrogen metabolism	7	reductase [NAD(P)H]	ec:1.7.1.2	2	map00910_20131216054321.png
Nitrogen metabolism	7	reductase (NADH)	ec:1.7.1.1	4	map00910_20131216054321.png
Nitrogen metabolism	7	reductase	ec:1.7.99.4	1	map00910_20131216054321.png
Flavonoid biosynthesis	10	3-dioxygenase	ec:1.14.11.9	4	map00941_20131216054326.png
Flavonoid biosynthesis	10	isomerase	ec:5.5.1.6	1	map00941_20131216054326.png
Flavonoid biosynthesis	10	synthase	ec:2.3.1.74	4	map00941_20131216054326.png
Flavonoid biosynthesis	10	3'-monooxygenase	ec:1.14.13.21	1	map00941_20131216054326.png
Pyruvate metabolism	9	carboxylase	ec:6.4.1.2	8	map00620_20131216054336.png
Pyruvate metabolism	9	lyase	ec:4.4.1.5	1	map00620_20131216054336.png
Phenylpropanoid biosynthesis	21	ammonia-lyase	ec:4.3.1.25	15	map00940_20131216054340.png

Phenylpropanoid biosynthesis	21	ammonia-lyase	ec:4.3.1.24	15	map00940_20131216054340.png
Phenylpropanoid biosynthesis	21	lactoperoxidase	ec:1.11.1.7	6	map00940_20131216054340.png
Lysine biosynthesis	12	N-succinyltransferase	ec:2.3.1.117	12	map00300_20131216054350.png
Arginine and proline metabolism	1	dehydrogenase [NAD(P)+]	ec:1.4.1.3	1	map00330_20131216054356.png
Arginine and proline metabolism	1	dehydrogenase	ec:1.4.1.2	1	map00330_20131216054356.png
Butanoate metabolism	4	hydratase	ec:4.2.1.17	4	map00650_20131216054402.png
Butanoate metabolism	4	epimerase	ec:5.1.2.3	3	map00650_20131216054402.png
Butanoate metabolism	4	dehydrogenase	ec:1.1.1.35	4	map00650_20131216054402.png
Benzoate degradation	4	hydratase	ec:4.2.1.17	4	map00362_20131216054407.png
Pentose and glucuronate interconversions	1	pectin demethoxylase	ec:3.1.1.11	1	map00040_20131216054416.png
Phenylalanine metabolism	25	hydratase	ec:4.2.1.17	4	map00360_20131216054422.png
Phenylalanine metabolism	25	ammonia-lyase	ec:4.3.1.25	15	map00360_20131216054422.png
Phenylalanine metabolism	25	ammonia-lyase	ec:4.3.1.24	15	map00360_20131216054422.png
Phenylalanine metabolism	25	lactoperoxidase	ec:1.11.1.7	6	map00360_20131216054422.png
Fatty acid degradation	4	hydratase	ec:4.2.1.17	4	map00071_20131216054428.png
Fatty acid degradation	4	epimerase	ec:5.1.2.3	3	map00071_20131216054428.png
Fatty acid degradation	4	dehydrogenase	ec:1.1.1.35	4	map00071_20131216054428.png
Fatty acid degradation	4	isomerase	ec:5.3.3.8	3	map00071_20131216054428.png
Streptomycin biosynthesis	1	synthase	ec:5.5.1.4	1	map00521_20131216054433.png
Amino sugar and nucleotide sugar metabolism	4	chitodextrinase	ec:3.2.1.14	4	map00520_20131216054438.png
Purine metabolism	16	phosphatase	ec:3.6.1.15	10	map00230_20131216054443.png

Purine metabolism	16	adenylpyrophosphatase	ec:3.6.1.3	6	map00230_20131216054443.png
Glycine, serine and threonine metabolism	5	dehydrogenase	ec:1.1.99.1	5	map00260_20131216054449.png
Carbon fixation in photosynthetic organisms	2	carboxylase	ec:4.1.1.39	2	map00710_20131216054453.png
Riboflavin metabolism	10	phosphatase	ec:3.1.3.2	10	map00740_20131216054458.png
Selenocompound metabolism	3	gamma-synthase	ec:2.5.1.48	1	map00450_20131216054504.png
Selenocompound metabolism	3	gamma-lyase	ec:4.4.1.11	3	map00450_20131216054504.png
Diterpenoid biosynthesis	5	13alpha-hydroxylase	ec:1.14.13.77	5	map00904_20131216054508.png
Glutathione metabolism	1	transferase	ec:2.5.1.18	1	map00480_20131216054516.png
Limonene and pinene degradation	4	hydratase	ec:4.2.1.17	4	map00903_20131216054521.png
Oxidative phosphorylation	11	oxidase	ec:1.9.3.1	2	map00190_20131216054526.png
Oxidative phosphorylation	11	dehydrogenase (quinone)	ec:1.6.99.5	1	map00190_20131216054526.png
Oxidative phosphorylation	11	reductase (H+-translocating)	ec:1.6.5.3	8	map00190_20131216054526.png
Terpenoid backbone biosynthesis	8	reductase (NADPH)	ec:1.1.1.34	3	map00900_20131216054534.png
Terpenoid backbone biosynthesis	8	diphosphate synthase	ec:2.5.1.29	1	map00900_20131216054534.png
Terpenoid backbone biosynthesis	8	Delta-isomerase	ec:5.3.3.2	4	map00900_20131216054534.png
Caprolactam degradation	4	hydratase	ec:4.2.1.17	4	map00930_20131216054538.png
Caprolactam degradation	4	dehydrogenase	ec:1.1.1.35	4	map00930_20131216054538.png
T cell receptor signaling pathway	5	protein-tyrosine kinase	ec:2.7.10.2	5	map04660_20131216054542.png
Propanoate metabolism	17	hydratase	ec:4.2.1.17	4	map00640_20131216054549.png
Propanoate metabolism	17	carboxylase	ec:6.4.1.2	8	map00640_20131216054549.png
Propanoate metabolism	17	ligase (ADP-forming)	ec:6.2.1.5	5	map00640_20131216054549.png

Tyrosine metabolism	6	oxidase	ec:1.10.3.1	6	map00350_20131216054554.png
Fatty acid elongation	4	hydratase	ec:4.2.1.17	4	map00062_20131216054558.png
Fatty acid elongation	4	dehydrogenase	ec:1.1.1.35	4	map00062_20131216054558.png
Fatty acid biosynthesis	10	carboxylase	ec:6.3.4.14	4	map00061_20131216054604.png
Fatty acid biosynthesis	10	carboxylase	ec:6.4.1.2	8	map00061_20131216054604.png
Tryptophan metabolism	4	hydratase	ec:4.2.1.17	4	map00380_20131216054609.png
Tryptophan metabolism	4	dehydrogenase	ec:1.1.1.35	4	map00380_20131216054609.png
Other glycan degradation	1	mannanase	ec:3.2.1.25	1	map00511_20131216054614.png
Aflatoxin biosynthesis	8	carboxylase	ec:6.4.1.2	8	map00254_20131216054618.png
Tetracycline biosynthesis	8	carboxylase	ec:6.4.1.2	8	map00253_20131216054622.png
Biosynthesis of unsaturated fatty acids	4	hydratase	ec:4.2.1.17	4	map01040_20131216054628.png
Alanine, aspartate and glutamate metabolism	2	synthase (NADH)	ec:1.4.1.14	1	map00250_20131216054634.png
Alanine, aspartate and glutamate metabolism	2	dehydrogenase [NAD(P)+]	ec:1.4.1.3	1	map00250_20131216054634.png
Alanine, aspartate and glutamate metabolism	2	dehydrogenase	ec:1.4.1.2	1	map00250_20131216054634.png
Geraniol degradation	4	hydratase	ec:4.2.1.17	4	map00281_20131216054638.png
Geraniol degradation	4	dehydrogenase	ec:1.1.1.35	4	map00281_20131216054638.png
Valine, leucine and isoleucine degradation	4	hydratase	ec:4.2.1.17	4	map00280_20131216054643.png
Valine, leucine and isoleucine degradation	4	dehydrogenase	ec:1.1.1.35	4	map00280_20131216054643.png

Thiamine metabolism	10	phosphatase	ec:3.6.1.15	10	map00730_20131216054651.png
beta-Alanine metabolism	4	hydratase	ec:4.2.1.17	4	map00410_20131216054655.png
Primary bile acid biosynthesis	4	dehydrogenase	ec:1.1.1.35	4	map00120_20131216054659.png
D-Glutamine and D-glutamate metabolism	1	dehydrogenase [NAD(P)+]	ec:1.4.1.3	1	map00471_20131216054706.png
Sulfur metabolism	1	gamma-synthase	ec:2.5.1.48	1	map00920_20131216054712.png
Drug metabolism - other enzymes	5	ali-esterase	ec:3.1.1.1	5	map00983_20131216054716.png
Drug metabolism - cytochrome P450	1	transferase	ec:2.5.1.18	1	map00982_20131216054721.png
Glyoxylate and dicarboxylate metabolism	2	carboxylase	ec:4.1.1.39	2	map00630_20131216054730.png
Isoquinoline alkaloid biosynthesis	6	oxidase	ec:1.10.3.1	6	map00950_20131216054735.png
Lysine degradation	4	hydratase	ec:4.2.1.17	4	map00310_20131216054742.png
Lysine degradation	4	dehydrogenase	ec:1.1.1.35	4	map00310_20131216054742.png
Ascorbate and aldarate metabolism	1	oxidase	ec:1.10.3.3	1	map00053_20131216054747.png
C5-Branched dibasic acid metabolism	5	ligase (ADP-forming)	ec:6.2.1.5	5	map00660_20131216054753.png
Metabolism of xenobiotics by cytochrome P450	1	transferase	ec:2.5.1.18	1	map00980_20131216054759.png
Galactose metabolism	4	invertase	ec:3.2.1.26	4	map00052_20131216054807.png
Citrate cycle (TCA cycle)	5	citrate synthase	ec:2.3.3.8	5	map00020_20131216054816.png
Citrate cycle (TCA cycle)	5	ligase (ADP-forming)	ec:6.2.1.5	5	map00020_20131216054816.png
Starch and sucrose metabolism	20	endo-1,3-beta-D-glucosidase	ec:3.2.1.39	15	map00500_20131216054823.png
Starch and sucrose metabolism	20	invertase	ec:3.2.1.26	4	map00500_20131216054823.png
Starch and sucrose metabolism	20	Pectin demethoxylase/methylesterase	ec:3.1.1.11	1	map00500_20131216054823.png

Inositol phosphate metabolism	1	synthase	ec:5.5.1.4	1	map00562_20131216054830.png
Glycerolipid metabolism	4	lipase	ec:3.1.1.3	4	map00561_20131216054836.png
alpha-Linolenic acid metabolism	8	hydratase	ec:4.2.1.17	4	map00592_20131216054840.png
alpha-Linolenic acid metabolism	8	13S-lipoxygenase	ec:1.13.11.12	4	map00592_20131216054840.png
Linoleic acid metabolism	4	13S-lipoxygenase	ec:1.13.11.12	4	map00591_20131216054844.png
Cysteine and methionine metabolism	4	gamma-synthase	ec:2.5.1.48	1	map00270_20131216054850.png
Cysteine and methionine metabolism	4	gamma-lyase	ec:4.4.1.11	3	map00270_20131216054850.png
Cysteine and methionine metabolism	4	oxidase	ec:1.14.17.4	1	map00270_20131216054850.png
Phenylalanine, tyrosine and tryptophan biosynthesis	4	dehydratase	ec:4.2.1.10	4	map00400_20131216054856.png
Phenylalanine, tyrosine and tryptophan biosynthesis	4	dehydrogenase	ec:1.1.1.25	4	map00400_20131216054856.png
Phenylalanine, tyrosine and tryptophan biosynthesis	4	dehydrogenase	ec:1.1.1.282	4	map00400_20131216054856.png
Carbon fixation pathways in prokaryotes	17	dehydrogenase	ec:1.1.1.35	4	map00720_20131216054900.png
Carbon fixation pathways in prokaryotes	17	citrate synthase	ec:2.3.3.8	5	map00720_20131216054900.png
Carbon fixation pathways in prokaryotes	17	carboxylase	ec:6.4.1.2	8	map00720_20131216054900.png
Carbon fixation pathways in prokaryotes	17	ligase (ADP-forming)	ec:6.2.1.5	5	map00720_20131216054900.png
Taurine and hypotaurine metabolism	1	dehydrogenase	ec:1.4.1.2	1	map00430_20131216054905.png

Table 4.4 KEGG pathways for down-regulated genes

Pathway	Seqs in Pathway	Enzyme	Enzyme ID	Seqs of Enzyme	Pathway Image
Retinol metabolism	5	monooxygenase	ec:1.14.14.1	5	map00830_20131031085514.png
Steroid hormone biosynthesis	5	monooxygenase	ec:1.14.14.1	5	map00140_20131031085520.png
Tryptophan metabolism	5	monooxygenase	ec:1.14.14.1	5	map00380_20131031085535.png
Linoleic acid metabolism	5	monooxygenase	ec:1.14.14.1	5	map00591_20131031085541.png
Arachidonic acid metabolism	5	monooxygenase	ec:1.14.14.1	5	map00590_20131031085558.png
Drug metabolism - cytochrome P450	5	monooxygenase	ec:1.14.14.1	5	map00982_20131031085607.png
Oxidative phosphorylation	1	NADH:ubiquinone reductase (H+-translocating)	ec:1.6.5.3	1	map00190_20131031085630.png
Metabolism of xenobiotics by cytochrome P450	5	monooxygenase	ec:1.14.14.1	5	map00980_20131031085709.png
Carbon fixation in photosynthetic organisms	3	ribulose-bisphosphate carboxylase/ribulose 1,5-bisphosphate carboxylase/oxygenase	ec:4.1.1.39	3	map00710_20131031085747.png
Glyoxylate and dicarboxylate metabolism	3	ribulose-bisphosphate carboxylase/ribulose 1,5-bisphosphate carboxylase/oxygenase	ec:4.1.1.39	3	map00630_20131031085958.png
Aminobenzoate degradation	5	monooxygenase	ec:1.14.14.1	5	map00627_20131031090029.png
Caffeine metabolism	5	monooxygenase	ec:1.14.14.1	5	map00232_20131031090042.png
Fatty acid degradation	5	monooxygenase	ec:1.14.14.1	5	map00071_20131031090103.png

Table 4.5 BLAST hits for up-regulated transcripts involved in host-pathogen interaction

Seq. Name	Seq. Description	Seq. Length	min. e-Value	mean Similarity
Signal Perception				
MDC002287.314.2	receptor-like protein 12-like	2735	0.0	64.1%
MDC007527.344.1	receptor-like protein	2950	0.0	88.05%
MDC021939.232.1	receptor-like protein 12-like	3028	0.0	67.45%
MDC024587.31.1	receptor-like protein	3052	0.0	80.9%
MDC005111.352.1	wall-associated receptor kinase-like 2-like	2656	0.0	85.8%
MDC005111.339.5	wall-associated receptor kinase-like 10-like	4298	0.0	83.75%
MDC005111.339.3	wall-associated kinase	1443	5.83033E-94	64.15%
MDC005111.339.2	wall-associated receptor kinase-like 10-like	3470	0.0	83.75%
MDC008279.241.1	probable lrr receptor-like serine threonine-protein kinase at4g08850-like	331	1.01074E-31	76.3%
MDC018545.561.5	probable receptor-like protein kinase at1g67000-like	3829	0.0	83.45%
MDC021678.122.1	probable lrr receptor-like serine threonine-protein kinase at3g47570-like	3200	0.0	69.95%
MDC004018.301.1	receptor-like protein kinase hsl1-like	3186	0.0	77.35%
MDC007170.566.1	probable leucine-rich repeat receptor-like protein kinase at1g35710-like	1562	1.79318E-116	55.25%
MDC020984.573.5	cysteine-rich receptor-like protein kinase 25-like	4432	9.75676E-98	64.2%
MDC020984.573.2	cysteine-rich receptor-like protein kinase 29-like	2563	4.14125E-81	63.0%
MDC020984.573.1	cysteine-rich receptor-like protein kinase 10-like	211	7.78395E-17	58.65%
MDC007092.798.7	probable lrr receptor-like serine threonine-protein kinase at1g56130-like	425	2.23882E-17	53.35%
MDC007092.798.4	leucine-rich repeat transmembrane protein kinase	786	5.77482E-63	82.45%
MDC007092.798.3	probable lrr receptor-like serine threonine-protein kinase at1g56130-like	969	4.15122E-22	75.7%
MDC007092.798.2	leucine-rich repeat transmembrane protein kinase	4682	1.04104E-46	76.45%
MDC009345.248.3	domain-containing receptor-like kinase	1042	3.43549E-17	61.25%
MDC002049.218.2	leucine-rich repeat receptor-like serine threonine tyrosine-protein kinase sobir1-like	3916	0.0	83.1%
MDC011518.548.8	probable lrr receptor-like serine threonine-protein kinase at3g47570-like	541	5.81304E-85	80.05%
MDC011518.548.7	probable lrr receptor-like serine threonine-protein kinase at3g47570-like	949	1.0492E-111	71.65%
MDC002426.446.2	probable lrr receptor-like serine threonine-protein kinase at4g08850-like	4589	2.46384E-97	77.05%
MDC002426.446.1	probable lrr receptor-like serine threonine-protein kinase at4g08850-like	1231	5.51702E-77	76.7%
MDC000839.494.5	receptor-like protein kinase	2054	3.63884E-16	47.6%
MDC000839.494.4	probable lrr receptor-like serine threonine-protein kinase at1g53430-like	2462	5.40775E-55	47.1%
MDC000839.494.1	probable lrr receptor-like serine threonine-protein kinase at1g53430-like	1382	3.4281E-32	70.45%
MDC019629.366.1	probable lrr receptor-like serine threonine-protein kinase at3g47570-like	3938	0.0	65.9%
MDC009645.284.1	probable receptor-like protein kinase at1g67000-like	1460	1.61789E-99	55.1%
MDC004079.537.2	receptor-like protein 12-like	3232	0.0	69.05%
MDC016149.194.4	serine threonine-protein kinase bri1-like 1	3442	0.0	80.75%
MDC001405.281.1	serine threonine-protein kinase bri1-like 1	3391	0.0	80.95%
MDC010317.647.3	brassinosteroid insensitive 1-associated receptor kinase 1	1107	1.69001E-35	66.85%

MDC022288.177.2	receptor-like protein kinase hsl1-like	814	1.18246E-117	82.85%
MDC006535.173.1	dual specificity protein kinase spla-like	1210	5.52146E-25	100.0%
MDC006536.318.1	kinesin-related protein 8-like isoform x2	3122	0.0	61.6%
MDC025597.72.3	kinesin-related protein 8-like isoform x2	4263	5.79366E-178	60.1%
Transporters				
MDC018794.69.2	sorbitol transporter	2563	0.0	87.55%
MDC008614.269.3	sorbitol transporter	4158	0.0	85.5%
MDC006067.135.4	sorbitol partial	2205	2.66981E-63	95.6%
MDC016275.388.1	abc transporter i family member 20-like	550	1.71062E-114	91.6%
MDC019951.202.2	abc transporter d family member chloroplastic-like	872	2.60522E-30	88.15%
MDC010152.224.1	abc transporter g family member 11-like	358	6.82948E-27	76.05%
MDC020082.224.2	pleiotropic drug resistance protein 2-like	6778	0.0	91.3%
MDC007204.286.2	phosphate transporter	659	1.12575E-133	94.35%
MDC001429.373.2	oligopeptide transporter 3-like	2488	1.33796E-92	93.9%
MDC010152.224.4	white-brown-complex abc transporter family	3675	0.0	81.45%
MDC012037.134.1	lysine histidine transporter-like 8-like	229	7.66257E-36	81.8%
MDC013395.260.3	probable plastidic glucose transporter 1-like	893	5.70938E-26	93.9%
MDC009483.39.2	lysine histidine transporter-like 8-like	227	1.01417E-19	93.5%
MDC000756.383.1	probable peptide nitrate transporter at5g62680-like	1068	7.0125E-124	75.65%
MDC013314.817.1	phosphate transporter isoform 1	1552	0.0	91.6%
MDC017657.246.1	blight-associated protein partial	1393	1.35925E-49	84.25%
Detoxification				
MDC012537.136.1	cytochrome p450 734a1-like	2289	6.07244E-142	74.4%
MDC003871.314.1	cytochrome p450 82a3-like	2044	1.03802E-138	72.0%
MDC015352.236.1	cytochrome p450	1106	5.82899E-170	83.85%
MDC011571.171.1	cytochrome p450	1186	0.0	75.75%
MDC006653.333.1	cytochrome p450 82a3-like	2401	7.07815E-138	71.35%
MDC000463.206.1	cytochrome p450	6209	1.6861E-136	78.6%
MDC007779.587.1	cytochrome p450 734a1-like	919	4.52783E-159	75.3%
MDC000838.123.1	cytochrome f	835	1.03479E-180	95.25%
MDC018770.180.1	apocytochrome b	405	2.57263E-87	98.75%
MDC000893.308.1	cytochrome p450 71a25-like	613	4.87174E-82	79.6%
MDC011028.370.2	cytochrome p450	1768	0.0	78.6%
MDC001715.363.1	cytochrome p450 71a1-like	826	2.5593E-175	71.65%
MDC013427.515.1	cytochrome p450 90a1-like	1665	1.65571E-53	89.55%
MDC006843.302.1	cytochrome p450 716b1-like	319	1.10648E-50	88.3%
MDC001204.808.1	glutathione s-transferase	1117	4.72536E-76	77.9%
MDC021421.104.2	glutathione s	328	4.70361E-58	88.75%
Secondary Metabolites				
MDC017443.214.1	chalcone synthase	1574	0.0	88.05%

MDC011812.165.1	chalcone synthase	743	9.26653E-154	98.5%
MDC009581.158.1	flavanone 3-dioxygenase-like	1956	4.70111E-60	87.15%
Resistance Proteins				
MDC015654.280.1	disease resistance family protein lrr family	1265	1.87638E-110	63.4%
MDC015654.424.1	disease resistance family protein lrr family	1600	8.37758E-133	61.75%
MDC008721.170.1	probable disease resistance protein at5g66900-like	375	4.93111E-31	65.75%
MDC020369.172.2	probable disease resistance protein at5g66900-like	1868	5.2072E-63	71.45%
MDC016420.630.4	nbs-lrr protein	785	4.23408E-43	59.1%
MDC000427.113.1	enhanced disease susceptibility 1	2408	0.0	70.75%
MDC022837.465.1	disease resistance protein rga3-like	1219	3.45626E-108	68.1%
MDC019950.237.1	disease resistance response protein 206-like	399	2.82335E-36	68.2%
MDC013853.361.1	disease resistance protein rpm1-like	2827	0.0	78.55%
Photosynthesis				
MDC013754.149.2	ribulose-bisphosphate carboxylase oxygenase large subunit	682	4.40302E-157	96.0%
MDC018698.251.1	photosystem ii cp43 chlorophyll partial	238	1.07474E-34	90.6%
MDC007463.318.1	photosystem i p700 apoprotein a2	932	0.0	99.95%
MDC015331.156.1	photosystem i p700 chlorophyll a apoprotein a1	790	3.85742E-166	98.0%
MDC006187.447.2	photosystem i p700 apoprotein a2	1004	0.0	99.95%
MDC002986.249.1	photosystem i p700 apoprotein partial	406	1.84807E-79	98.65%
Transcription factors				
MDC033588.6.4	probable wrky transcription factor 51-like	3606	4.31046E-23	86.5%
MDC017895.316.1	probable wrky transcription factor 33-like	1974	0.0	74.75%
MDC035519.7.1	probable wrky transcription factor 70-like	328	1.01308E-22	58.0%
MDC007698.491.1	wrky transcription	449	5.40556E-43	57.45%
MDC015642.226.1	wrky transcription factor 6-like	158	7.98041E-20	79.5%
MDC001475.298.3	myb 12 transcription factor	3671	7.34482E-41	62.7%
MDC011101.167.1	zinc finger	1287	1.41101E-38	93.75%
Defence				
MDC013032.460.1	beta-glucanase	946	0.0	87.1%
MDC013032.473.2	beta-glucanase	1231	0.0	87.9%
MDC015282.544.1	beta-glucanase	973	0.0	83.1%
MDC030146.5.1	beta-glucanase	913	0.0	88.35%
MDC009167.350.1	thaumatin-like protein	691	2.91189E-138	86.4%
MDC011784.298.2	class iv chitinase	2738	5.78221E-78	86.3%
MDC011784.298.1	class iv chitinase	616	1.32266E-85	86.85%
Cell wall modification				
MDC010817.271.5	beta-glucosidase 24-like	2448	5.19178E-44	75.45%

MDC009345.248.5	plant invertase pectin methylesterase inhibitor	1790	2.99961E-16	63.85%
MDC007696.347.1	plant invertase pectin methylesterase inhibitor	2224	5.41223E-13	62.88%
MDC011414.254.4	phenylalanine ammonia-lyase	2990	0.0	95.65%
MDC022151.329.4	phenylalanine ammonia-lyase	3419	0.0	96.1%
MDC022151.340.2	phenylalanine ammonia-lyase	3107	0.0	96.2%
MDC017169.202.1	phenylalanine ammonia-lyase	3174	0.0	94.75%
MDC015907.151.1	peroxidase 3-like	472	6.55516E-77	80.9%

Table 6.1 BLAST hits for transcriptome sequences four dpi

Seq. Name	*Functional Description*	*Seq. Length*	*min. e Value*	*mean Similarity*
Signal perception				
MDC000726.370.2	probable serine threonine-protein kinase at4g35230-like isoform 2	181	5.66943E-25	91.7%
MDC010746.224.2	serine threonine-protein kinase rio1-like	1237	4.2585E-29	60.15%
MDC000226.496.5	reticuline oxidase-like	1588	0.0	83.75%
MDC000726.370.1	receptor protein	622	9.87509E-43	99.55%
MDC022760.144.1	leucine-rich repeat containing	641	3.31016E-100	66.55%
MDC011524.743.1	leucine-rich repeat receptor protein kinase exs-like	3523	0.0	79.55%
Defence proteins				
MDC002485.209.2	glucan endo-beta-glucosidase 11 isoform 3	580	1.7987E-66	91.75%
Transport				
MDC003519.375.1	abc transporter b family member 8-like	1855	1.64501E-70	91.25%
MDC022358.366.1	atp binding protein	589	1.8513E-110	86.65%
MDC022358.211.1	atp binding protein	481	4.62228E-88	81.3%
Detox				
MDC026986.10.3	cytochrome p450 like tbp	238	1.79342E-42	83.85%
MDC012440.297.5	cytochrome p450 71a25-like	2881	8.65159E-32	79.75%
MDC012901.193.1	cytochrome p450 like tbp	552	2.25697E-27	89.4%
MDC008223.286.1	cytochrome p450 like tbp	317	4.94169E-52	90.05%
MDC004040.483.1	cytochrome p450 like tbp	134	8.6702E-22	91.36%
MDC006042.67.1	cytochrome p450 like tbp	199	4.52778E-27	87.68%
MDC005345.248.2	thioredoxin-like 1- chloroplastic-like	1003	5.2755E-96	75.7%
Disease Resistant				
MDC017187.315.1	syntaxin t-snare family	1888	4.40529E-81	58.73%
MDC004759.111.5	syntaxin of plants	220	5.41642E-37	90.85%
MDC001689.403.1	disease resistance gene	622	4.03985E-90	72.6%
Transcription				
MDC011591.183.2	probable wrky transcription factor 70-like	325	5.22574E-15	58.2%
MDC004759.111.1	transcription initiation factor tfiid subunit 6-like	2533	1.2209E-16	89.9%
MDC012233.162.1	trihelix transcription factor gt-3b-like	211	6.71864E-31	89.4%
MDC023703.33.1	zinc finger protein	540	1.64981E-45	61.65%
MDC018645.239.1	ring finger protein	720	3.52149E-58	70.75%

Hormone synthesis				
MDC019559.196.2	gibberellin 3-beta-dioxygenase 4-like	172	3.9927E-30	91.3%
Protein synthesis				
MDC001225.168.3	eukaryotic initiation factor 4a-3-like	247	1.18267E-16	75.75%
Protein destination and storage				
MDC005345.248.1	chaperone protein dnaj chloroplastic-like	418	5.26352E-39	88.1%
Response to abiotic stress				
MDC016220.263.1	galactinol synthase	1128	1.71498E-38	92.65%

Table 6.2 BLAST hits for protein sequences four dpi

Seq. Name	*Functional Description*	*Seq. Length*	*min. e Value*	*mean Similarity*
Redox				
Spot1	Abscisic acid stress ripening protein homolog	233	8.9423E-10	57.5%
Spot7	Chloroplast heat shock protein 70	737	0.0	85.4%
Spot11	Thioredoxin	180	5.85448E-80	80.35%
Energy				
Spot2	Triosephosphate isomerase	254	1.83741E-148	92.05%
Spot3	Triosephosphate isomerase	254	1.39823E-148	92.05%
Spot4	Ribulose-bisphosphate carboxylase oxygenase large subunit	226	1.54227E-156	95.8%
Spot12	Ribulose bisphosphate carboxylase oxygenase activase chloroplast	492	0.0	85.8%
Spot13	Oxygen-evolving enhancer protein chloroplast	332	0.0	94.3%
Disease/Defense				
Spot5	Mlp-like protein	153	1.08178E-62	76.7%

Appendix B

```
**********************************************************************

>library(limma)
>library(edgeR)

#Read everything in and merge the technical replicates
timepoint <- factor(c(0,2,4,8,12,0,2,4,8,12))
conds <- factor(c(rep("untreated", 5), rep("treated",5)),
levels=c("untreated", "treated"))
Data2 <- read.table("merged_counts.txt", header=T, row.names=1)
Data2$untreated1 <- Data2$untreated1 + Data2$untreated1.1
Data2$untreated4 <- Data2$untreated4 + Data2$untreated4.1
Data2$untreated5 <- Data2$untreated5 + Data2$untreated5.1
Data2$treated1 <- Data2$treated1 + Data2$treated1.1 +
Data2$treated1.2
Data2$treated2 <- Data2$treated2 + Data2$treated2.1 +
Data2$treated2.2
Data2$treated4 <- Data2$treated4 + Data2$treated4.1 +
Data2$treated4.2
Data2 <- Data2[,-c(2,6,8,10,11,13,14,17,18)]
design <- model.matrix(~conds+timepoint)

nf <- calcNormFactors(Data2)
v <- voom(Data2, design=design, plot=F, lib.sizes=colSums(Data)*nf)
plotMDS(v, top=50, labels=names(Data),
col=ifelse(conds==unique(conds)[1],"blue","red"),
gene.selection="common")

#fit the model
vf <- lmFit(v, design=design)
vf <- eBayes(vf)

#Do we have DE genes?
summary(decideTests(vf))
vennDiagram(decideTests(vf)[,-1])

#Extract the DE genes and make an intersection
tt <- topTable(vf, coef="condstreated", n=1000000)
DEtreatment <- subset(row.names(tt), tt$adj.P.Val < 0.1)
DEtimepoints <- c()
for(i in c(2:5)) {
    tt <- topTable(vf, coef=i, n=1000000)
    DEtimepoints <- c(DEtimepoints, subset(row.names(tt),
tt$adj.P.Val < 0.1))
}
DEtimepoints <- unique(DEtimepoints)

This section answered the question on how to extract DE's for
timepoint4
> tt5 <- topTable(vf2, coef="timepoint4", n=400000)
> USE5 <- which(tt5$adj.P.Val < 0.05)
```

```
> length(which(tt5$adj.P.Val < 0.05))
Extract genes from tt5 using the following:
USE <- which(tt5$adj.P.Val<0.05)
tt5[USE,]
write.table(tt5[USE5,], file="Timepoint4.xls", sep="\t", col.names =
NA)
**********************************************************************

**********************************************************************
The following scripts and commands were used to extract the genes
from the apple genome and submitted to Blast2GO.

perl fetch.MDP.with.RLOC_v2.pl apple_genome.gtf 398_DE2.txt >
398_2.RLOC

grep -f 398_DE.txt apple_genome.gtf > miniature_annotation_398.gff
For up-regulated genes:
copy up-regulated RLOC's in Excel to Word (paste special,
unformatted text, save as plain text, M_Dos. pico file, save as Mac
format in Terminal (Mac application).
grep -f up-regulated.txt apple_genome.gtf >
miniature_annotation_up.gff

perl gtf2bed.pl miniature_annot_up.gff > minUp.bed

Use this command:
perl gtf2bed.pl miniature_annotation_398.gff > min_2.bed to get only
1164 RLOC/MDPs which correspond to 398.RLOC file

Open the .bed file in Excel, copy first 3 columns in Excel file to
Word (paste special, unformatted text, save as plain text, M_Dos.
pico file save as Mac format in Terminal (Mac application).

Convert apple_genome to one_liner:
perl fasta.one.line.pl apple_genome.fa_copy > apple_one_line.fasta

To extract sequences from apple_genome (only 1164 in this case)
perl substr.pl Header_list_miniature_annotation2.txt
apple_one_line.fasta > 1164.fasta


Open min_162.bed in Excel, copy 3 columns to Texteditor file. pico
from Mac format to Dos format.

perl substr.pl Header_162.txt apple_one_line.fasta > 162.fasta

For all the ID's (398)
perl substr.pl Header.min_2.txt apple_one_line.fasta > ALL2.fasta

To extract ID's from topTable
perl fetch.by.IDfromtextfile.pl 398_DE.txt toptable2.txt > 398logFC

Use this script to re-name the repeats in the ALL2.fasta file for
Blast2GO input:
```

```
perl reconcile.repeated.ids.pl 398.txt > 398.2.txt
```

****

Appendix C

convert.py script

****

```
#!/usr/bin/env python

import csv

import re

import sys


f = csv.reader(open("apple_genome.gff3", "r"), dialect="excel-tab")

of = open("apple_genome.cleaned.gff3", "w")


offset = 0

current_contig = ""

for line in f :

    #The version or other similar lines can simply be copied over

    if(line[0][0] == "#") :

        of.write("%s\n" % line[0])

        continue


    #skip any chromosome or cluster lines

    if(line[2] == "chromosome" or line[2] == "cluster") :

        continue


    #If we've hit a new contig, store the name and the associated
offset

    if(line[2] == "contig") :
```

```
        offset = int(line[3]) - 1

        matches = re.search("ID=([\w\.]+);", line[8])

        if(matches == None) :

            sys.exit(line)

        current_contig = matches.group(1)



    of.write("%s\t%s\t%s\t%i\t%i\t%s\t%s\t%s\t%s\n" %
(current_contig, line[1], line[2], int(line[3]) - offset,
int(line[4]) - offset, line[5], line[6], line[7], line[8]))



of.close()
**********************************************************************

merge.py script
**********************************************************************
#!/usr/bin/python
#########################################################
#   example.py - a program to ....                      #
#                                                       #
#                                                       #
# Purpose: merge count tables                           #
#                                                       #
# Usage: python merge_tables guide_file                 #
#########################################################
#looks for text guide_file
#that contains files to be merged with column headers (space
separated)
#ie
#file1.counts untreated1
#file2.counts untreated2
#file3.counts treated1
#file4.counts treated2
#this will generate a tab separated table like this
#
#gene untreated1 untreated2 treated1 treated2
#gene1 0 0 0 0
#gene2 1 0 11 10
#.......
#################################################
import sys
try:
      infile = open(sys.argv[1])
except IndexError:
      print "No guide file provided"
      sys.exit()
```

```
#make dict of genes with list of counts
#list is ordered so treatments will be preserved.
#genes = {'gene1':[1,2,3,4]}
#header keeps track of treatment order, will be as read from config
col_header = []
genes = {}

outfile = open('merged_counts.txt','w')

for line in infile:
      filename,header = line.strip().split(' ')
      try:
            data_f = open(filename)
      except IOError:
            print "%s can't be found?"%filename
            sys.exit()

      col_header.append(header)

      #read file and add gene and counts to the dict
      for line in data_f:
            gene,count = line.strip().split('\t')
            if gene not in genes:
                  genes[gene] = [count]
            else:
                  genes[gene].append(count)
      #important to close file
      data_f.close()

infile.close()

outfile.write('gene\t'+'\t'.join(col_header)+'\n')

for gene in genes:
      data = genes[gene]
      #make sure each treatment has a count for this gene
      #this should catch most errors
      try:
            assert len(data) == len(col_header)
      except AssertionError:
            print "one of the treatment or genes is missing or extra"
            print "data, found the problem here:"
            print gene,data
            print "while %s columns of treatments
given" %len(col_header)
            sys.exit()

      out_data = gene+'\t'+'\t'.join(data)+'\n'
      outfile.write(out_data)

outfile.close()
print "Merged table is 'merged_counts.txt'"

**********************************************************************
```

fetch.MDP.with.RLOC_v2.pl

```
**********************************************************************
#!/usr/bin/perl -w

#AUTHOR Mbandi S.K (ksmbandi@sanbi.ac.za)
#South African National Bioinformatics Institute

my $gtf = shift || die "\nusage: perl fetch.MDP.with.RLOC.pl
GTF.file RLOC.file\n\n";
my $rloc = shift || die "\nusage: perl fetch.MDP.with.RLOC.pl
GTF.file RLOC.file\n\n";;
my $temp = "${gtf}.tab";
my (%hash, %tab);
open (GTF, $gtf);
open (TAB, "> $temp");
while (<GTF>){
      $_ =~ m/.+(MDP\d{10}).+(RLOC_\d{8})/;
      my $transcript = $1; #print $transcript."\n";
      my $locus = $2; #print $$locus."\n";
      $hash{$transcript} = $locus;
}
foreach my $key ( keys %hash ){
  print TAB $hash{$key}."\t".$key."\n";
}
close GTF;
close TAB;
open (TAB, $temp);
while(<TAB>){
      chomp $_;
      my @data = split("\t", $_);
      $tab{$data[1]} = $data[0];
}
close TAB;

open (RLOC, $rloc);
while (<RLOC>) {
        chomp $_;
        foreach my $key ( keys %tab ){
            if ($_ eq $tab{$key}){
                  print $_."\t".$key."\n";
                  #print $key."\n";
                 }
        }
}
close RLOC;

**********************************************************************
```

gtf2bed.pl

```
**********************************************************************
#!/usr/bin/perl

# Copyright (c) 2011 Erik Aronesty (erik@q32.com)
#
use Data::Dumper;

$in = shift @ARGV;

open IN, ($in =~ /\.gz$/ ? "gunzip -c $in" : $in =~ /\.zip$/ ?
"unzip -p $in" : "$in");
while (<IN>) {
$gff = 2 if /^##gff-version 2/;
$gff = 3 if /^##gff-version 3/;
next if /^#/ && $gff;

s/\s+$//;
# 0-chr 1-src 2-feat 3-beg 4-end 5-scor 6-dir 7-fram 8-attr
my @f = split /\t/;
if ($gff) {
# most ver 2's stick gene names in the id field
($id) = $f[8]=~ /\bID="([^"]+)"/;
# most ver 3's stick unquoted names in the name field
($id) = $f[8]=~ /\bName=([^";]+)/ if !$id && $gff == 3;
} else {
($id) = $f[8]=~ /transcript_id "([^"]+)"/;
}

next unless $id && $f[0];

if ($f[2] eq 'exon') {
die "no position at exon on line $." if ! $f[3];
# gff3 puts :\d in exons sometimes
$id =~ s/:\d+$// if $gff == 3;
push @{$exons{$id}}, \@f;
# save lowest start
$trans{$id} = \@f if !$trans{$id};
} elsif ($f[2] eq 'start_codon') {
#optional, output codon start/stop as "thick" region in bed
$sc{$id}->[0] = $f[3];
} elsif ($f[2] eq 'CDS') {
#optional, output codon start/stop as "thick" region in bed
push @{$cds{$id}}, \@f;
# save lowest start
$cdx{$id} = \@f if !$cdx{$id};
} elsif ($f[2] eq 'stop_codon') {
```

```
$sc{$id}->[1] = $f[4];
} elsif ($f[2] eq 'miRNA' ) {
$trans{$id} = \@f if !$trans{$id};
push @{$exons{$id}}, \@f;
}
}

for $id (
# sort by chr then pos
sort {
$trans{$a}->[0] eq $trans{$b}->[0] ?
$trans{$a}->[3] <=> $trans{$b}->[3] :
$trans{$a}->[0] cmp $trans{$b}->[0]
} (keys(%trans)) ) {
my ($chr, undef, undef, undef, undef, undef, $dir, undef, $attr,
undef, $cds, $cde) = @{$trans{$id}};
my ($cds, $cde);
($cds, $cde) = @{$sc{$id}} if $sc{$id};

# sort by pos
my @ex = sort {
$a->[3] <=> $b->[3]
} @{$exons{$id}};

my $beg = $ex[0][3];
my $end = $ex[-1][4];

if ($dir eq '-') {
# swap
$tmp=$cds;
$cds=$cde;
$cde=$tmp;
$cds -= 2 if $cds;
$cde += 2 if $cde;
}

# not specified, just use exons
$cds = $beg if !$cds;
$cde = $end if !$cde;

# adjust start for bed
--$beg; --$cds;

my $exn = @ex; # exon count
my $exst = join ",", map {$_->[3]-$beg-1} @ex; # exon start
my $exsz = join ",", map {$_->[4]-$_->[3]+1} @ex; # exon size

# added an extra comma to make it look exactly like ucsc's beds
print
"$chr\t$beg\t$end\t$id\t0\t$dir\t$cds\t$cde\t0\t$exn\t$exsz,\t$exst,
\n";
}
```

```
close IN;
**********************************************************************
```

```
fasta.one.line.pl
**********************************************************************
#!/usr/bin/perl

#AUTHOR Mbandi S.K (ksmbandi@sanbi.ac.za)
#South African National Bioinformatics Institute

use strict;
use warnings;

die "usage fasta.one.line.pl fasta\n" unless (@ARGV==1);
my $fasta = $ARGV[0];
open (IF, $fasta) || die "$fasta does not exist\n";
my $count = 0;
my $sep = 1;
while (my $line = <IF>){
        chomp $line;
        if ($line =~ m/^>/){
                $count+=1;
                if ($sep != $count){
                print "\n";
                }
                print $line."\n";
        }
        else{
                unless($line =~ />/){
                        my $seq .= $line; print $seq;
                }
        }
} if ($count > 1){print "\n";}
close IF
**********************************************************************
```



```
substr.pl
**********************************************************************

#! /usr/bin/perl -w

#AUTHOR Mbandi S.K (ksmbandi@sanbi.ac.za)
#University of the Western Cape

#Will isolate a substring of a sequence if the sequence id, start
and stop coordinates are specified
```

```
#example: isolate the region that represents a domain or gene
segment
#########################################################################

if (@ARGV != 2){
      print STDERR "\nusage: substr.pl <file containing ids, start,
stop> <genome>\n\n";
      exit(1);
}

my $ids = $ARGV[0];
my $seqfile = $ARGV[1];

my $start="";
my $stop="";
open(F, $ids);

while (my $entry = <F>) {
      chomp $entry;
      my @line = split("\t", $entry);
      my $vi = $line[0];
      ($start, $stop)= ($line[1]-1, $line[2]);
      my $bytes = ($line[2] - $line[1]) + 1;
      &get_dat($vi, $start, $bytes);
}
close(F);

sub get_dat{
my ($vi,$offset, $size)= @_;
open (G, $seqfile) || die "$seqfile does not exist\n";
while(my $id=<G>){
      $seq=<G>;
      $id =~ s/>//g;
      chomp ($id,$seq); #print $id."\t".$vi."\n"; exit;
      if ($vi =~ /$id/){
                  my $substr = substr($seq, $offset, $size);
                  print ">".$id."\n".$substr."\n";

            }
      }
close G;
}

**********************************************************************


fetch.by.IDfromtextfile.pl
**********************************************************************
#!/usr/bin/env perl

#AUTHOR Mbandi S.K (ksmbandi@sanbi.ac.za)
#University of the Western Cape
#########################################################################
```

```
use strict;
use warnings;

open FILE1, "< test.txt" or die "could not open file1\n";
my $keyRef;
while (<FILE1>) {
   chomp;
   $keyRef->{$_} = 1;
}
close FILE1;

open FILE2, "< 398_DE.txt" or die "could not open file2\n";
while (<FILE2>) {
    chomp;
    my ($testKey, $label, $count) = split("\t", $_);
    if (defined $keyRef->{$testKey}) {
        print STDOUT "$_\n";
    }
}
close FILE2;
**********************************************************************


reconcile.repeated.ids.pl
**********************************************************************
#!/bin/perl -w
#AUTHOR Mbandi S.K (ksmbandi@sanbi.ac.za)
#University of the Western Cape


use Bio::SeqIO;
if (@ARGV != 1){
      print STDERR "\nusage: reconcile.repeated.ids.pl <file
format>\n\n";
      exit(1);
}
use Bio::SeqIO;
my $file = shift;
my $cnt = 0;
$seqio_obj = Bio::SeqIO->new(-file => "$file", -format => "fasta");
#Use the next_seq method of Bio::SeqIO to ctreate a sequence object
while ($seq_obj = $seqio_obj->next_seq()){
      #print ">".$seq_obj->display_id."\n".$seq_obj->seq."\n";
      my $id = $seq_obj->display_id;
      my $seq = $seq_obj->seq;
      if (defined $hash{$id}){
            $hash{$id} = "$hash{$id}\t$seq";
      }
      else{
            $hash{$id} = $seq;
      }
}
while (($key, $value) = each %hash){
      my @entry = split("\t", $value);
```

```
      my $num_seq = scalar @entry;
      foreach (@entry){
            print ">$key.$num_seq\n$_\n";
            $num_seq--;
      }
}
```

****