UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Transcriptome profiling of potato plant stems challenged with *Pectobacterium carotovorum* subsp. *brasiliense* and elucidation of the role of small RNAs in *Pectobacterium* survival mechanisms

By

Stanford Kwenda

Submitted in partial fulfilment of the requirements for the degree

Doctor of Philosophy (Microbiology)

In the Faculty of Natural & Agricultural Sciences

University of Pretoria

Pretoria

2016, 14 October

SUPERVISORY COMMITTEE

Prof L. N. Moleleki

Department of Microbiology and Plant Pathology
Forestry and Agricultural Biotechnology Institute (FABI)
University of Pretoria
Lunnon Road, Hatfield
Pretoria
0028
South Africa

Prof P. R. J. Birch

Division of Plant Sciences
College of Life Sciences,
University of Dundee (James Hutton Institute)
Errol Road, Invergowrie
Dundee, DD25DA
Scotland, UK.

DECLARATION

I, Stanford Kwenda, declare that the thesis/dissertation, which I hereby submit for the degree PhD Microbiology at the University of Pretoria, is my own work and has not previously been submitted by me for a degree at this or any other tertiary institution.

SIGNATURE: …...……..

DATE: ……..

ACKNOWLEDGEMENTS

I would like to deeply thank the following people for their helpful contributions to the research work in this thesis:

- My supervisor, Prof L. N. Moleleki, for the professional guidance and unwavering support throughout the duration of my PhD studies. Thank you for your patience and understanding; and allowing me to 'think outside the box'. Most importantly I am grateful for everything I learnt under your mentorship.
- My co-supervisor, Prof P.R.J. Birch, I am grateful for the time you took to critically review my work. Thank you for the insightful comments and advice.
- My lab mates, thank you guys for the friendship and support, and for bearing with me throughout. I want to give special mention to Dr Ramesh Aadi Moolam and Tshepiso Motlolometsi for the assistance you gave with the RT-PCRs and gels. I greatly appreciate the assistance rendered. Special thanks to Collins Tanui for being always ready to provide helpful hands.
- To my wife, Heather, thank you for the patience and unwavering support throughout. You have been a pillar of support, and a constant source of joy, happiness and love. I love you so much.
- To my family, my heartfelt thankfulness to you guys for being there for me always, and the constant support and encouragement throughout the duration of my studies. Thank you mum and Kudzi for the constant prayers. I greatly appreciate you all.
- Above all, I want to give thanks to my Lord and savior, Jesus Christ, from whom I draw my knowledge, wisdom and strength.

I would also like to acknowledge the following institutions and funding bodies:

- National Research Foundation (NRF)
- Forestry and Agricultural Biotechnology Institute (FABI)
- University of Pretoria Postgraduate Research Bursary

SUMMARY

Pectobacterium carotovorum subsp. *brasiliense*, a necrotrophic phytopathogen belonging to the soft rot *Enterobacteriaceae* (SRE) family is responsible for causing tuber soft rot and blackleg diseases of stems in potato plants. In recent years, *P. c. brasiliense*, has emerged as a soft rot pathogen of significance, potentially threatening potato production globally. To date, *P. c. brasiliense* is the most aggressive soft rot phytopathogen isolated from potato in South Africa. Currently effective chemical control measures are unavailable once soft rot pathogens have established disease in potato plants and/or harvested tubers. Therefore, this study sought to determine the molecular basis of quantitative resistance in potato stems challenged with *P. c. brasiliense*. In addition, this thesis explores some of the regulatory mechanisms important in the adaptation of *Pectobacterium* species to harsh nutrient-deficient environments such as plant xylem vessels. Determining the activated defense responses in potato stems is key in deciphering potential control approaches against pectobacteria as these soft rot pathogens colonize vascular tissues during infection of plants. Currently, no transcriptome-wide studies have been applied in the *P. c. brasiliense* and potato stem interaction to understand inducible defense responses within potato stems.

In chapter 2, by implementing a time-course RNA-seq analysis, our study revealed important signaling pathways suggested to contribute to the potato defense transcriptome against *P. c. brasiliense* infection. Comparison of transcriptomes between a susceptible potato cultivar (*Solanum tuberosum* cv Valor) and tolerant cultivar (*S. tuberosum* cv BP1) following *P. c. brasiliense* inoculation revealed that the MAPK signaling cascades and ethylene hormonal pathway are central to potato defense responses against this pathogen. Specifically, genes encoding MPK3 protein kinase, and MKS1; ethylene biosynthetic and signaling pathways such as *ACC*, *ERF2* and *EIN3* genes were up-regulated in the tolerant cultivar within the time-course. Furthermore, expression of downstream defense-related genes was enhanced in *S. tuberosum* cv BP1, including transcription factors such WRKY33, MYB83, and several

ethylene-responsive binding factors (ERFs); as well as various secondary wall biosynthetic genes for lignification and cellulose biosynthesis, for example, *IRX9* and *CESA8*, respectively.

In chapter 3, a bioinformatics analysis using strand-specific RNA sequencing allowed the identification of 1113 potato long intergenic noncoding RNA (lincRNAs) from stem tissues. Long noncoding RNAs (lncRNAs) have been implicated in diverse regulatory roles in eukaryotes. Recently, defense-related lncRNAs have been identified in Arabidopsis and wheat. In this thesis we identified 559 potato lincRNAs that were differentially expressed (DE) in both cultivars compared to mock-inoculated controls, following inoculation by *P. c. brasiliense*. Furthermore, co-expression analysis associated 17 of these lincRNAs with 12 potato defense-related genes. These results suggest that lincRNAs possibly have functional roles in potato defence responses. Future work will focus on characterization of these lincRNAs in order to understand their specific functional roles, particularly in potato defense mechanisms.

In chapter 4, regarding potential regulatory mechanisms employed by *Pectobacterium* species during survival under nutrient-limiting conditions, we described 137 sRNA transcripts in *P. atrosepticum* genome. About 62% of the identified sRNAs are conserved within the SRE. Furthermore, 68 sRNAs were differentially expressed when comparing *P. atrosepticum* cells under growth-promoting and starvation conditions; with 47 sRNAs up-regulated under nutrient-deficient conditions. Thus, since many starvation-induced sRNAs were identified, these findings highlighted that sRNAs play key roles in adaptive responses in the genus *Pectobacterium*.

THESIS OUTPUTS

Scientific publications:

Kwenda, S., Gorshkov, V., Ramesh, A. M., Naidoo, S., Rubagotti, E., Birch, P. R. J., Moleleki, L. N. (2016) Discovery and profiling of small RNAs responsive to stress conditions in the plant pathogen *Pectobacterium atrosepticum***.** *BMC Genomics* **17**(1) 10.1186/s12864-016-2376-0.

Kwenda, S., Birch, P. R. J., Moleleki, L. N. (2016) Genome-wide identification of potato long intergenic noncoding RNAs responsive to *Pectobacterium carotovorum* subspecies *brasiliense* infection. *BMC Genomics* **17**(614)10.1186/s12864-016-2967-9.

Kwenda, S., Birch, P. R. J., Moleleki, L. N. (2016) RNA-seq profiling reveals defense responses in a tolerant potato cultivar to stem infection by *Pectobacterium carotovorum* ssp. *brasiliense*. *Frontiers in Plant Science*. **7**:1905. doi: 10.3389/fpls.2016.01905.

Congresses:

National:

- SA Genetics Society (SAGS) & SA Society for Bioinformatics (SASBi) Joint Congress
  23-26 Sep 2014 (Pretoria, South Africa)
  Oral Presentation:
  Kwenda, S., Gorshkov, V., Ramesh, A. M., Naidoo, S., Rubagotti, E., Birch, P. R. J., Moleleki, L. N. (2014) Strand-specific RNA-seq and computational prediction of small RNAs in the plant pathogen *Pectobacterium atrosepticum*

- Genomics Research Institute (GRI) Seminar Day
  31 Oct 2014 (University of Pretoria)
  Oral Presentation:
  Kwenda, S., Gorshkov, V., Ramesh, A. M., Naidoo, S., Rubagotti, E., Birch, P. R. J., Moleleki, L. N. (2014) Discovery of small RNAs in the plant pathogen *Pectobacterium atrosepticum*

- Genomics Research Institute (GRI) Seminar Day
  31 Oct 2015 (University of Pretoria)
  Oral Presentation:
  Kwenda, S., Birch, P. R. J., Moleleki, L. N. (2016) Identification and characterization of long noncoding RNA in potato responsive to *Pectobacterium carotovorum* subsp. *brasiliense* infection.

International:

- 18th International Plant Protection Congress (IPPC 2015)
  24-27 Aug 2015 (Berlin, Germany)
  Poster Presentation:
  Kwenda, S., Mosina, G, Birch, P. R. J., Moleleki, L. N. (2015) Transcriptome profiling of potato cultivars under *Pectobacterium carotovorum* subspecies *brasiliense* challenge.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

3` UTR:	3` untranslated regions
5` UTR:	5` untranslated regions
ACO:	ACC oxidase
ACS:	ACC synthase
AHLs:	Acyl homoserine lactones
asRNAs:	Antisense RNAs
BAK1:	Brassinosteroid insensitive 1-associated kinase 1
BIK1:	Botrys induced kinase 1
BLAST:	Basic Local Alignment Search Tool
BOS1:	Botrytis susceptible 1
bp:	base pairs (bp)
BP:	Biological process
CC:	Cellular component
CC domain:	Coiled-coil domain
CDS:	coding DNA sequence
CFU:	colony forming units
DAMPs:	Damage-Associated Molecular Patterns
DE:	Differentially expressed
DEGs:	Differentially expressed genes
EFR:	EF-Tu Receptor
EF-Tu:	Elongation factor Tu
EPS:	Extracellular polysaccharides
ERF1:	Ethylene response factor 1
ERF104:	Ethylene response factor 104
ET:	Ethylene
ETI:	Effector-triggered immunity
FDR:	False discovery rate
FLS2:	FLAGELLIN SENSING 2
GO:	Gene ontology

Hpi:	Hours post inoculation
HR:	Hypersensitive response
IGR:	Intergenic regions
JA:	Jasmonic acid
JA-ET:	jasmonic/ethylene pathways
LincRNAs:	long intergenic noncoding RNAs
lncRNAs:	long noncoding RNAs;
LRR-RLK:	Leucine-rich repeat receptor-like kinase
MAPK:	Mitogen-activated protein kinase
MCP:	Methyl-accepting chemotaxis
MF:	Molecular function
miRNA:	micro RNAs
mRNA:	messenger RNA
NBS-LRR:	Nucleotide binding site leucine rich repeat
ncRNAs:	noncoding RNAs
NGS:	next generation sequencing
OGs:	Oligogalacturionides
ORFs:	Open reading frames
PAD3:	Phytoalexin deficient 3
PAMPs:	Pathogen-Associated Molecular Patterns
Pcb1692:	*Pectobacterium carotovorum* subsp. *brasiliense* 1692
PCWDEs:	Plant cell wall degrading enzymes
PE reads:	Paired-end reads
PRRs:	pattern recognition receptors
PTI:	pattern triggered immunity
QS:	Quorum sensing
R:	Resistance gene(s)
RITs:	Rho-independent terminators
RLCK:	Receptor-like cytoplasmic kinase
RLK:	Receptor-like kinase
RNA-seq:	RNA sequencing

ROS:	Reactive oxygen species
RPKM:	Fragments per kilo bases of exons for per million mapped reads
RT-qPCR:	Reverse-transcription quantitative PCR
SA:	Salicylic acid
SERK:	Somatic embryogenesis receptor kinase
SIPK:	Salicylic acid-induced protein kinase
siRNAs:	short interfering RNAs
snoRNAs:	small nucleolar RNAs
snRNA:	small nuclear RNAs
SRE:	Soft rot enterobacteria/ Soft rot *Enterobacteriaceae*
sRNA:	small RNA
ssRNA-seq:	Strand-specific RNA sequencing
T2SS:	Type II secretion system
T3SS:	Type III secretion system
T4SS:	Type IV secretion system
T5SS:	Type V secretion system
T6SS:	Type VI secretion system
TARs:	Transcriptionally active regions
TIR:	Toll and interleukin-1-like receptor domain
TPP:	Thiamine pyrophosphate
TSS:	Transcriptional start sites
TTSS:	Type III secretion systems
WAK1:	Wall-associated kinase 1
WAK2:	Wall-associated kinase 2
WIPK:	Wounding-induced protein kinase

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
Literature review

1.1 General Introduction

The genus *Pectobacterium* which belongs to the bacterial family *Enterobacteriaceae* infects a variety of plants worldwide resulting in soft rot and blackleg diseases. *Pectobacterium* species have been isolated from both monocotyledonous and dicotyledonous plants (Glasner et al., 2008). *Pectobacterium* species typically exhibit a necrotrophic feeding mode during infection and they employ an array of cell wall degrading enzymes to macerate host plant tissue (Toth and Birch, 2005). These species cause soft rot of potato tubers, stem rot (which typically develops from the top of the plant and progresses downwards to the stem base), wilt and blackleg (wet slimy black rot lesion which spreads up the stems from rotting mother tubers) diseases in potato plants (Czajkowski et al., 2011).

Several species belonging to this genus appear to have a relatively limited host-range. Among these, are *P. atrosepticum*, *P. wasabiae*, *P. betavasculorum* and *P. carotovorum* subsp. *odoriferum*, and their main hosts include potato, horseradish, sugar beet, and chicory, respectively. However, *P. carotovorum* subsp. *carotovorum* and *P. carotovorum* subsp. *brasiliense* are broad-host range phytopathogens, in particular, *P. carotovorum* subsp. *carotovorum* has the widest host range worldwide (Pérombelon, 2002, Glasner et al., 2008). Even though the host range of *P. carotovorum* subsp. *brasiliense* is still not fully understood, it is emerging as an aggressive broad host range soft rot pathogen of global significance (Marquez-Villavicencio et al., 2011). To date, most of the pectobacteria have been isolated from potato worldwide with the exception of *P. betavasculorum* (Charkowski, 2015). Evidently, potato is the main host of *Pectobacterium*, and these pathogens have a major economic impact on potato production. Indeed the genus *Pectobacterium* is complex, comprising strains with diverse characteristics, consequently this presents challenges in devising ways to control soft rot related diseases.

Potato (*Solanum tuberosum*) is produced globally and is the fourth most important food crop in the world behind rice (*Oryza sativa*), wheat (*Triticum* aestivum), and maize (*Zea* mays) (http://faostat.fao.org/site/339/default.aspx). The majority of potato cultivars are highly heterozygous, autotetraploid (2n =4x = 48), and belong to the *Solanaceae* family, including tomato, pepper, and eggplant (Consortium, 2011). Over the past two decades, potato production has steadily increased globally, with marked increases in the developing countries (Birch et al., 2012). In South Africa, for example, approximately 2 million tons of potato are produced annually, and the average yields have been steadily increasing to about 40 t/ha (http://nbsystems.co.za/potato/index_3.htm). These current global potato production trends can be attributed to improved agronomic practices and use of high yielding cultivars. Indeed, potato has become central to global food security especially within the developing world as

its tubers provide an important dietary source of starch, vitamins, proteins and antioxidants (Burlingame et al., 2009, Birch et al., 2012). However, most potato cultivars are highly susceptible to several pests and pathogens. This is likely due to the narrow genetic base of cultivated potato as a result of acute inbreeding depression over the decades (Consortium, 2011). Furthermore, application of pesticides is only useful against those pathogens that can be controlled. This excludes bacterial diseases such as soft rot and wilt, against which there is no commercially available chemical control measures (Birch et al., 2012, Czajkowski et al., 2011, Charkowski, 2015). Thus, one of the major focuses is to investigate novel ways for pathogen control including searches for the molecular basis of potato innate defenses and identification of the potato gene complement which confers disease resistance. Owing to the availability of the high-quality whole genome sequence of potato (*S. tuberosum* group Phureja DM1-3 516 R44) (Consortium, 2011), it is now possible to perform genome-wide functional analyses in order to decipher the underlying mechanisms and key genes in potato physiology, development, and its interaction with- and resistance to various pathogens. The potato genome size is suggested to be approximately 844 Mb, with 39,031 predicted protein-coding genes (Consortium, 2011).

1.2 Plant Innate Immunity

Plants as sessile organisms constantly come into contact with potential microbial pathogens in the soil and in the air throughout their life cycles. Thus, they can be infected by a large number of these phytopathogens which results in diseases on important crop plants in terms of food and feed. Consequently plant diseases, if unmanaged, can result in heavy economic losses and food shortages, thus, threatening food security globally. Over the decades, breeding crop plants for resistance has proven a highly effective and environmentally friendly approach towards disease prevention. Breeding for resistance has been achieved by exploitation of plant resistance genes, which are components of the broader innate immunity of plants (Dodds and Rathjen, 2010).

Plant innate immunity is inducible and serves as the first line of defense against infection (Asai et al., 2002). Thus, in order to fend off and inhibit multiplication of invading pathogens, plants naturally depend on innate immune responses of each cell, activated by perceived signals at infection sites (Jones and Dangl, 2006). The plant innate immunity is a two-branched system comprising two layers of pathogen detection strategies. Conserved and slowly evolving microbial features such as fungal chitin and bacterial flagellin, collectively known as Microbe-Associated Molecular Patterns or Pathogen-Associated Molecular Patterns (PAMPs) are perceived by receptor proteins, referred to as pattern recognition receptors (PRRs), localized on the surfaces of the host cells, resulting in pattern triggered immunity (PTI) (Dodds and

Rathjen, 2010, Henry et al., 2013). Besides microbial features, plant PRRs can also recognize endogenous molecules referred to as Damage-Associated Molecular Patterns (DAMPs) due to wounding during pathogen invasion. DAMPs can be cuticular or cell wall fragments. Likewise, recognition of DAMPs initiates the PTI immune response.

Invading pathogens have, however, evolved the ability to attenuate PTI by preventing perception by the extracellular surface receptors or by releasing effector molecules directly into the cytoplasm. In turn, these effectors alter resistance signaling or the resulting resistance responses (Chisholm et al., 2006). Therefore, the second recognition strategy mediated by intracellular immune receptors, involves perception of specific virulence factors called effectors. This perception leads to robust effector-triggered immunity (ETI) (Dodds and Rathjen, 2010, Henry et al., 2013). In general, PTI and ETI commonly induce similar downstream immune responses, however, ETI is faster and more amplified than PTI, often resulting in a localized hypersensitive response (HR) or cell death at the infection sites, indicating resistance (Dodds and Rathjen, 2010, Jones and Dangl, 2006). Thus, ETI is generally effective against biotrophs and hemibitrophs, whereas, PTI is widely effective against most pathogens, particularly, non-adapted necrotrophic pathogens (Dodds and Rathjen, 2010, Jones and Dangl, 2006, Henry et al., 2013). Overall, the innate immune system of plants is very efficient, conferring most plants with resistance to most pathogens (Zipfel, 2008).

1.2.1 Pattern-Triggered Immunity

PTI serves as the first layer of the plant innate defense response initiated upon recognition of PAMPs by PRRs and PTI usually stops infection before the pathogen is established in the plant (Chisholm et al., 2006). Intracellular responses related to PTI include early responses such as induction of ion fluxes, reactive oxygen species (ROS) production, mitogen-activated protein kinase (MAPK) signaling, ethylene biosynthesis, transcriptional activation of defense genes and callose deposition for reinforcement of cell walls as a late response (Zipfel, 2008). Plants recognize diverse PAMPS which may include lipids, proteins, carbohydrates and small molecules such as acyl homoserine lactones, key signals in bacterial quorum sensing (Boller and Felix, 2009). However, PRRs for perception of some of these PAMPs are yet to be fully characterized.

Currently, the best understood plant PRR, is the leucine-rich repeat receptor-like kinase (LRR-RLK), FLAGELLIN SENSING 2 (FLS2), in Arabidopsis which recognizes bacterial flagellin and directly binds to the 22 amino acid epitope, flg22 in the N-terminus of flagellin (Gómez-Gómez and Boller, 2000, Zipfel, 2008). Functional homologs of FLS2 have been found in

various plants including some belonging to the *Solanaceae* family such as tomato (Robatzek et al., 2007). Perception of flg22 results in the assembly of an active signaling complex between FLS2 and BAK1 (Brassinosteroid insensitive 1-associated kinase 1), a co-receptor which has a central role in PTI (Dodds and Rathjen, 2010). BAK1 is another LRR-RLK that is required by most known PRRs for functioning. BAK1 is part of the somatic embryogenesis receptor kinase (SERK) family. Additionally, BAK1 is involved in sensing DAMPs and various PAMPs (Dodds and Rathjen, 2010, Henry et al., 2013). However, *bak1* mutants are transiently sensitive to FLS2 implying that other SERK members could also be playing yet to be characterized roles in immune signaling (Zipfel, 2008). Since BAK1 plays a central role in PTI regulation, it is potentially targeted by phytopathogen virulence effectors (Shan et al., 2008). Elongation factor Tu (EF-Tu) is another abundantly available bacterial protein, recognized as a PAMP in Arabidopsis. Specifically, the N-acetylated 18 amino acid peptide, elf18, within EF-Tu is the primary signature for recognition (Kunze et al., 2004). The plant PRR for EF-Tu recognition is EF-Tu Receptor (EFR), an LRR-RLK within the same subfamily as FLS2 (Zipfel et al., 2006). As with FLS2, EFR also associates with BAK1, upon stimulation with EF-Tu.

Collectively, signaling components downstream of plant PRRs appear conserved in diverse plant families. Specifically, multiple PTI signaling pathways likely activate defenses via convergence at overlapping MAPK cascades and transcription factors (Zipfel, 2008, Asai et al., 2002). For instance BAK1, positively regulates both FLS2 and EFR. Additionally BAK1 is possibly important in responses to necrotrophs (Zipfel, 2008). BAK1 seems not to have direct roles in PAMP perception, even though it rapidly forms complexes with PRRs such as FLS2 following elicitation (Dodds and Rathjen, 2010). An additional regulator of the FLS2 complex, is Botrys induced kinase 1 (BIK1), a receptor-like cytoplasmic kinase (RLCK), lacking the extracellular LRR domain. In Arabidopsis, BIK1 has a major part in PTI signaling, integrating responses, not only downstream of FLS2, but for EFR and CERK1 receptors as well. Thus, BIK1 brings together responses from various PRRs but its function is thought to be independent of MAPK signaling. Instead, BIK1 activation is suggested to be ethylene (ET)-signaling dependent (Mengiste, 2012). Furthermore, BIK1 potentially regulates the phosphorylation of the FLS2-BAK1 complex by interacting with FLS2 and BAK1 before stimulation and dissociating from the complex after stimulation (Dodds and Rathjen, 2010). However, BIK1 function in plant immunity is still not fully characterized, owing to contrasting results in which *bik1* mutant Arabidopsis plants appear more resistant to *Pseudomonas syringae* infection in comparison to wild-type Arabidopsis plants due to over production of the salicylic acid hormone, however, these mutants were susceptible to inoculation by *Botrytis cinerea*, a necrotrophic fungus (Lu et al., 2010, Veronese et al., 2006). Thus, there is need for more studies to fully decipher the roles of BIK1 in plant immunity responses. To date, the

known and fully characterized PRRs in plants that recognize bacterial PAMPs are FLS2 and EFR (Zipfel, 2008).

1.2.2 DAMP-Triggered Immunity

The cuticle and cell wall in plants are dynamic structures functioning as barriers inhibiting infection, however, lytic cell wall degrading enzymes mainly secreted by necrotrophic pathogens result in degradation products produced by the action of these exoenzymes such as cell wall fragments serving as sources of DAMPS. Typically DAMPs appear in the apoplast, and like PAMPs, they are detected by infected plants as danger signals leading to stimulation of PTI innate immunity. Therefore, in plants, cell wall fragments are classic examples of DAMPs, and their derivatives such as oligogalacturionides (OGs) induce PTI (Boller and Felix, 2009). Perception of these modified-self or endogenous elicitors such as DAMPs triggers defenses which enhance basal immune responses, mainly PTI (Davidsson et al., 2013). Collectively, PTI is centered on PRR-mediated perception of PAMPs and DAMPs (Boller and Felix, 2009). The recognition of DAMPs occurs within the vicinity of wounding and pathogen invasion sites. In Arabidopsis, the perception of OGs derived from cell walls is mediated by the PRR, wall-associated kinase 1 (WAK1) and WAK2 (Brutus et al., 2010). WAK1 is an RLK, and homologs of WAK1 are present in other plant families. During DAMP recognition, the extracellular domain of WAK1 binds with OGs. Similar to flg22, perception of OGs induces typical PTI responses including lignification of cell wall, phytoalexin accumulation, oxidative burst and hormone biosynthesis. As such, OG-PTI in plants is more important against necrotrophs but contributes minimally to plant responses in biotrophic interactions. For instance, increased resistance to *B. cinerea* was noted in Arabidopsis following overexpression of WAK1 (Mengiste, 2012). Thus, wall-associated kinases (WAKs) serve as PRRs, monitoring pectin disruption by plant cell wall degrading enzymes (PCWDEs) (Mengiste, 2012).

Additional DAMPs associated with plant immune responses include the small peptides *At*Pep1 and Pep2, identified in Arabidopsis leaves (Huffaker et al., 2006, Huffaker and Ryan, 2007). These small peptides represent DAMPs for wounding and stress and they contribute to defense signal amplification and cell to cell communication in plants (Mengiste, 2012, Boller and Felix, 2009). In Arabidopsis, the LRR-RLKs, PEPR1 and PEPR2 were identified as receptors for *At*Pep1 and Pep2, respectively (Boller and Felix, 2009, Mengiste, 2012, Yamaguchi et al., 2006). Strikingly, the pairing of Pep1/PEPR1 resembles flg22/FLS2 and elf18/EFR ligand-receptor pairs. Finally, like the PAMP receptors (FLS2 and EFR), PEPR1, PEPR2 and WAK1, also form complexes with BAK1, thus linking the perceived DAMP signals to PTI defense responses (Boller and Felix, 2009, Mengiste, 2012).

1.2.3 Effector-Triggered Immunity

Pathogenic effectors are detected by host plant resistance (*R*) proteins intracellularly, leading to ETI mediated disease resistance that often leads to localized HR cell death at infection sites. *R* proteins are intracellular immune receptors encoding polymorphic nucleotide binding site (NBS)-LRR protein domains (Jones and Dangl, 2006). Accordingly, NBS-LRR-mediated immune defense responses are effective against biotrophic and hemibiotrophic pathogens which require and derive nutrients from living host tissue. However, ETI-disease resistance is less effective against necrotrophs which aggressively kill tissue during host infection (Jones and Dangl, 2006, Wang et al., 2014).

NBS-LRR domain containing proteins can perceive corresponding effectors by means of direct physical association or indirectly via effector-mediated alterations of proteins interacting with *R* genes (Elmore et al., 2011). Individual phytopathogenic bacterial strains can deliver multiple effectors (around 15-30) into plant cells via type III secretion systems (TTSS). Microbial pathogenic effector molecules promote virulence usually through mimicry or inhibition of host cellular functions. Some effectors possibly promote nutrient leakage, others perform structural roles or aid in pathogen dispersal, while the majority may directly interfere with various components of PTI or suppress ETI-mediated disease resistance (Dodds and Rathjen, 2010, Jones and Dangl, 2006). In the Arabidopsis genome, approximately 125 *R* genes encode NBS-LRR proteins. Based on their N-terminus, plant NBS-LRR proteins fall into two groups, possessing either the coiled-coil (CC) domain or Toll and interleukin-1-like receptor (TIR) domain (Meyers et al., 2003).

Evidently, plants recognize PAMPs, DAMPs and effector molecules as danger signals inducing the defense response signaling (Boller and Felix, 2009). Downstream gene expression signals in PTI and ETI are largely the same, suggesting overall similarity in the responses but varying in magnitude (Tao et al., 2003). Thus, many cellular events are attributed to both PTI and ETI, including rapid calcium ion influx, ROS, activation of MAPK cascades, gene expression reprogramming, cell wall callosic deposition at attempted infection sites and HR cell death (Dodds and Rathjen, 2010)

1.2.4 MAPKs and downstream activation of defense responses

The transduction of sensed stress signals into appropriate downstream responses is an important step for the survival and adaptation of plants (Pitzschke et al., 2009). Plant kinase cascades belonging to the mitogen-activated protein kinase (MAPK) group have a crucial role in signaling diverse biotic and abiotic stresses. In addition, the MAPK cascade-mediated signal transduction step is remarkably critical in resistance establishment against pathogens

(Pitzschke et al., 2009). Activation of MAPK signaling cascades is part of the earliest signaling events when plants sense the invading pathogens (Meng and Zhang, 2013). Stimulation of MAPKs occurs through a phosphorelay mechanism in which perceived signal from upstream receptors is relayed via MAPKKK (MAPK kinase kinase or MEKK), onto MAPKK (MAPK kinase or MEK) and finally MAPKs, thus, linking perceived cues to downstream targets (Pitzschke et al., 2009, Meng and Zhang, 2013). Accordingly, MAPK cascades comprise three kinases and are important signaling modules functioning in converting sensed signals into cellular responses.

MAPK cascade modules are highly conserved within the eukaryotes, however, compared to yeast and animals, classes of MAPK, MAPKK, and MAPKKK families are more expanded in plants. Based on sequence homology, the Arabidopsis genome putatively comprises over 80 MAPKKKs, 10 MAPKKs and 20 MAPKs suggesting the complexity of MAPK cascades (Ichimura et al., 2002). Although several MAPKs have been identified, the best characterized are MPK3/MPK6 and MPK4. MPK3/MPK6 are functionally redundant and are activated following abiotic and/or biotic stress stimuli as well as in growth and development. Importantly, MPK3/MPK6 positively regulate defense responses in plants. In addition MPK4 and MPK11 are also part of the plant defense mechanisms against pathogen infection (Tena et al., 2011, Pedley and Martin, 2005, Cristina et al., 2010, Bethke et al., 2012). However, identities of most MAPKKKs and MAPKKs are mostly unclear (Asai et al., 2002). Together, MAPK cascades regulate plant development and signal responses to various stresses including wounding, drought, salinity, temperature, ROS and pathogen attack (Pitzschke et al., 2009, Ichimura et al., 2002, Meng and Zhang, 2013).

1.2.4.1 Plant MAPK signaling in defense responses

There is still a dearth of knowledge about the signaling components operating directly downstream of the activation of extracellular PRRs and intracellular NBS-LRR proteins leading to kinase-cascade signaling activation. Components of these intermediary signaling pathways continue to be elusive, with limited success from genetic screens in identifying these missing links in plant immunity signal transduction. However, the MAPK signaling pathway is now known to be involved in defense responses in both PTI and ETI and much attention and understanding has been gained in MAPK signaling over the years (Dodds and Rathjen, 2010).

1.2.4.2 Pattern-Triggered Immunity MAPK signaling

MAPK signaling cascades have a central role in the PTI pathway by transduction of signal from PRRs to immune response components downstream (Chisholm et al., 2006, Cristina et al., 2010). Several plant PRRs are implicated in MAPK signaling stimulation after PAMP

recognition (Meng and Zhang, 2013). In Arabidopsis, MAPK cascades acting downstream of flg22 and elf18 perception have been characterized. Perception of flg22 activates two MAPK cascades in Arabidopsis.

The first MAPK signaling pathway downstream of FLS2 consists of MAPKKKs (redundant orthologs of MEKK1), MKK4/MKK5 (two functionally redundant MAPKKs) and MPK3/MPK6 (partially redundant MAPKs) (Asai et al., 2002). However, genetic studies revealed that *mekk1* mutants show normal MPK3/MPK6 activation upon treatment with flg22, thus, concluding that MEKK1 is not responsible for MKK4/MKK5 phosphorylation and activation. Instead MEKK1 is involved in the activation of the second flg22-activated MAPK cascade, comprising MEKK1, MKK1/MKK2 (redundant MAPKKs) and MPK4 (Ichimura et al., 2002, Suarez-Rodriguez et al., 2007). The MPK4 cascade results in positive regulation of basal defenses (Meng and Zhang, 2013). Stimulation of MPK3/MPK6 positively regulates defense responses by stimulation of WRKY-type transcription factors (possibly functionally redundant with WRKY33) which culminates in the expression of defense genes. Furthermore, MPK3 and/or MPK6 are triggered by other PAMPs such as the fungal elicitor chitin (Boller and Felix, 2009, Miya et al., 2007). Additionally, DAMP perception, also triggers the MPK3/MPK6 signaling cascade in OG-activated defense signaling. For instance, in Arabidopsis, MPK3/MPK6 are stimulated by OGs.

Collectively, different plant PRRs regulate stimulation of very similar sets of genes, indicating that MAPKs represent the point of convergence in PTI signaling. In PTI signaling, MAPK cascades control and regulate diverse downstream defense responses including modulation of synthesis and/ or signaling of defense hormones, stimulation of transcription factors, activation of antimicrobial metabolites such as phytoalexins, control of ROS and other defense genes (Meng and Zhang, 2013, Tena et al., 2011).

1.2.4.3 Effector-Triggered Immunity MAPK signaling

MAPK cascades were shown as fundamental signal transduction components of ETI in tobacco and tomato studies (Pedley and Martin, 2005, Oh and Martin, 2011). In transgenic tobacco cells, SIPK (salicylic acid-induced protein kinase) and WIPK (wounding-induced protein kinase), orthologs of Arabidopsis MPK3 and MPK6, respectively, were triggered by the fungal Avr9 effector and by Tobacco mosaic virus (TMV), expressing the cognate Cf-9 resistance gene and in an N-resistance-gene dependent manner (Meng and Zhang, 2013). Additionally, the tomato Pto-mediated pathway revealed activation of MAPKs in ETI defense response (Oh and Martin, 2011). Pto is an intracellular protein kinase receptor which recognizes the bacterial effectors AvrPto or AvrPtoB, leading to activation of the NBS-LRR

resistance protein Prf, which then triggers ETI responses (Oh and Martin, 2011). The MAPK cascade composed of MAPKKKα, MEK2, and SIPK/WIPK positively regulates Pto-triggered ETI in tobacco (del Pozo et al., 2004). Furthermore, MPK3/MPK6 have been implicated in Arabidopsis ETI responses (Tsuda et al., 2009). The MAPK cascade consisting of MKK7 and MPK3/MPK6 leads to salicylic acid (SA) accumulation, constitutive expression of *PR-1*, increased basal resistance to biotrophs, and activation of SAR (Meng and Zhang, 2013, Zhang et al., 2007). Typically, ETI results in HR-related cell death, a process linked to MAPK activation, production of ROS, and accumulation of SA (Greenberg and Yao, 2004, Coll et al., 2011).

1.2.5 Plant immune responses against necrotrophs (post MAPK-signal transduction)

Plants respond to pathogen attack by activating various defense response mechanisms including antimicrobial metabolites and proteins, HR cell death, and production of lignin and callose for cell wall reinforcements (Glazebrook, 2005). Which defense mechanisms are activated is dependent on the type of invading pathogen. According to their modes of nutrition, phytopathogens can be grouped broadly into biotrophs (pathogens which can coevolve with and derive their nutrients from living plant host tissues), necrotrophs (those that extract nutrients from dying or dead cells killed during colonization or prior) and hemibiotrophs, which display an early biotrophic lifestyle followed by the necrotrophic feeding mode. However, the switching point in their feeding lifestyle varies among hemibiotrophs. Typically, necrotrophs secrete various phytotoxic compounds and PCWDEs to induce necrosis and leakage of nutrients (Mengiste, 2012).

Generally ETI-mediated defenses are associated with modulation of SA-dependent signaling pathway, mainly important against biotrophic and hemibiotrophic pathogens, and lead to regulation of certain pathogenesis-related genes which contribute to resistance (Glazebrook, 2005). The jasmonic acid (JA) and ethylene (ET) hormone signaling pathways are mainly involved in defenses against necrotrophs (Dodds and Rathjen, 2010, Glazebrook, 2005). Despite the considerable differences in the resulting gene expression from these pathways, there is substantial overlap between the SA and JA-ET pathways (Dodds and Rathjen, 2010). SA, ET and JA plant hormones are key signaling components in plant defense. Thus, induction of these hormones during pathogen invasion results in activation of various defense-related genes. However, central to the induction of plant hormonal signaling in plant immunity are MAPK cascades, which are involved in regulation of both defense hormone biosynthetic pathways and responses downstream of these hormonal signaling pathways (Meng and Zhang, 2013).

In plants, elevated levels of SA promotes resistance to hemibiotrophs but enhances susceptibility to necrotrophs. SA is mutually antagonistic with ET-JA signaling in plant immune responses (Meng and Zhang, 2013, Glazebrook, 2005). Downstream components of SA include NPR1 (a major regulator of SA-mediated defense responses), TGA2, TGA5 and TGA6 (redundant transcription factors) (Meng and Zhang, 2013). Plants produce elevated levels of ethylene when under pathogen attack. Furthermore, ET-mediated signaling in immune responses is crucial for resistance to necrotrophs and ET is suggested to be central to PTI defense responses against various pathogens including necrotrophs (Meng and Zhang, 2013, Mengiste, 2012, Lai and Mengiste, 2013). The ethylene biosynthetic pathway comprises two enzymatic steps which are catalyzed by ACC synthase (ACS) and ACC oxidase (ACO) to form ethylene. Components of ET include EIN2, EIN3 (which regulates FLS2 and BIK1 in PTI and is a major regulator of ethylene responses), accumulation of secondary metabolites and callose in flg22-PTI, and ET-mediated cell wall modifications (Mengiste, 2012). In summary, ET regulates various components of PTI signaling, including genes in the PAMP receptor complex, activation of MAPKs and BIK1, various transcription factors and expression of defense-related genes.

The JA-signaling pathway is mainly involved in plant defenses against insect pests and necrotrophic pathogens (AbuQamar et al., 2008). JA and ET exhibit synergistic regulation of many defense response genes and resistance in Arabidopsis, but tend to be functionally independent in tomato (Mengiste, 2012, AbuQamar et al., 2008). Furthermore, in tomato, *SPR2* (required for JA synthesis), *JAI1* (JA receptor in tomato), *ACX1* and *DEF1* are involved in basal resistance to the necrotrophic fungus, *B. cinerea* (Mengiste, 2012). Additional components of JA-signaling include *JAR1* which encodes a JA-amino synthetase which forms conjugates between JA and various amino acids. JA-mediated plant immune responses to nectrophic pathogens include protease inhibitor regulation and biosynthesis of secondary metabolites such as the antimicrobial flavonoid, anthocyanin and downstream induction of *ERF1* (ethylene response factor), MYC2 and *RAP2.6* (AP2 family) transcription factors in Arabidopsis (El Oirdi et al., 2011, Shan et al., 2009, Glazebrook, 2005, Mengiste, 2012).

Finally, the outcome of PTI in plants, is the activation of defense responses that result in inhibition of microbial colonization. Consequently, late PTI responses include cell wall modifications through secondary cell wall biosynthesis and callose deposition. Callose deposition probably occurs around 16 h after inoculation as was observed on Arabidopsis leaves treated with flg22 (Boller and Felix, 2009). Plant cells deposit callose between the cell wall and cell membrane within the vicinity of the invading pathogen, thus, callose deposits (or papillae) help to block cellular penetration at infection sites (Freeman and Beattie, 2008). In

Arabidopsis, papillae are synthesized by callose synthases (e.g. Protein POWDERY MILDEW RESISTANT 4), which is primarily responsible for defense-induced papillae production (callose deposition) (Malinovsky et al., 2015).

1.2.5.1 Transcription factors

In addition to regulating defense genes through controlling hormone signaling, MAPKs, can phosphorylate downstream transcription factors directly, thereby, activating defense genes. Thus, interaction of MAPKs with transcription factors and stimulation of defense responses after the release of transcription factors is a common mechanism in plant immunity (Lai and Mengiste, 2013). Accordingly, transcription factors connect upstream perception and early MAPK signaling events to downstream gene expression. Several transcription factor families are associated with plant immunity to necrotrophic pathogens including ERFs, WRKYs and MYBs.

The Arabidopsis defensin genes, *PDF1.2a* and *PDF1.2b* are activated by ERF104 (an ethylene response factor) after interaction and phosphorylation by MPK6 in an ET-dependent manner (Bethke et al., 2009). Additionally, ERF proteins modulate ET and JA responses and ERFs such as ORA59, act by integrating JA and ET signals in Arabidopsis, thus playing important roles in PTI resistance. Therefore, interaction of ERF transcription factor genes with MPK3/MPK6 suggests that they are components of the PTI pathway.

Several MYB transcription factors mediate host response by either promoting or suppressing plant resistance through different mechanisms. MYB51 and MYB122 are implicated in activation of genes involved in indole glucosinolate biosynthesis. Ultimately, this leads to downstream synthesis of secondary metabolites such as indole glucosinolates and camalexin, which enhance resistance against necrotrophic fungi (Mengiste, 2012, Frerigmann et al., 2015). The R2R3 MYB (Botrytis susceptible 1; BOS1) also plays significant roles against necrotrophic pathogens by restricting necrosis triggered during pathogen attack (Mengiste et al., 2003). Additionally, MYB transcription factors such as MYB46/MYB83 are master regulators of secondary wall formation in Arabidopsis, which directly regulate the expression of genes involved in lignin, cellulose and hemicellulose biosynthesis, including among others, *SND1*, *CESA4*, CESA7, *CESA8*, MYB58, MYB56 and MYB63 (McCarthy et al., 2009, Ko et al., 2014). Arabidopsis *myb46* mutants tend to have enhanced resistance to invading pathogens, thus, MYB46 could be thought of as a contributor to plant susceptibility to necrotrophs by suppressing cell wall defenses (Ramírez et al., 2011). However, the disease resistance phenotype observed in the *myb46* mutant could be partly due to constitutively activated plant immune responses, for example, accumulation of antimicrobial metabolites

was elevated much higher in this mutant in comparison to wild-type *Arabidopsis* plants (Hernández-Blanco et al., 2007, Miedes et al., 2015).

Among the transcription factors associated with plant immune responses, WRKY33, plays a significant function in resistance against necrotrophic pathogens in Arabidopsis. Furthermore, in Arabidopsis, WRKY33 has been shown to regulate camalexin biosynthesis and expression of the autophagy gene, *ATG18a* (Mengiste, 2012). WRKY33 is regulated by MPK3/MPK6 and MPK4 and its DNA binding activity is stimulated when it interacts with SIGMA FACTOR BINDING PROTEINs (SIB1 and SIB2). In *Arabidopsis*, resistance to *B. cinerea* is enhanced when SIB1 is overexpressed but compromised in *sib1 sib2* double mutant (Lai and Mengiste, 2013).

1.2.5.2 Phytoalexin accumulation

In response to attack and infection by various pathogenic microorganisms, plants produce low molecular weight secondary antimicrobial compounds known as phytoalexins, which are significant components of the defense arsenal of plants. Different classes of phytoalexins are produced by different plant families (Ahuja et al., 2012). Plants produce many phytoalexins, most of which are derived from the phenylpropanoid, isoprenoid, fatty acid/polyketide or alkaloid pathways (Dixon, 2001). Generally, related plant families tend to utilize similar chemical compounds for defense. For example, *Leguminosae* typically produce isoflavonoids, and *Solanaceae* employ sesquiterpenes such as capsidiol. However, some secondary metabolite classes such as phenylpropanoid derivatives, are used across plant taxa for defensive functions (Dixon, 2001). Most phytoalexins exhibit relatively broad spectrum activity and their specificity tends to be determined by the invading pathogen's mode of nutrition and enzymatic machinery (Dixon, 2001). Signal-transduction pathways orchestrate the accumulation of inducible antimicrobial secondary metabolites following recognition of pathogens by host receptors (Jones and Dangl, 2006).

Phytoalexins can be functionally defined as compounds synthesized *de novo* after infection or abiotic stress (Dixon, 2001, VanEtten et al., 1994, Ahuja et al., 2012). In Arabidopsis, camalexin, is the major phyotalexin and its production can be induced by various biotrophic and necrotrophic pathogens (Mao et al., 2011, Ahuja et al., 2012, Meng and Zhang, 2013). The biosynthetic pathways of most phytoalexins are largely unknown, including the regulation of their induction by abiotic or biotic stress, however, camalexin regulation in Arabidopsis has been relatively well studied (Ahuja et al., 2012). Activation of the camalexin biosynthetic pathway, involves MPK3/MPK6 and MPK4 signaling and regulation by WRKY33. WRKY33 binds to the promoter region of *PAD3* (Phytoalexin deficient 3), directly activating camalexin

biosynthetic gene expression (Mao et al., 2011). In tobacco, SIPK and WIPK (homologs of MPK3/MPK6 module) phosphorylate and activate a WRKY33-like NbWRKY8 transcription factor, leading to induction of key genes such as *3-hydroxy-3-methylglutaryl CoA reductase 2* (*HMGR2*) in the production of isoprenoid phytoalexins (Ishihama et al., 2011). *In vitro* test assessing the toxicity of camalexin to bacterial or fungal pathogens revealed that camalexin antimicrobial activity includes disruption of bacterial membranes and fungal cell membrane damage, and induction of fungal apoptotic-like programmed cell death, thus, restricting spread of lesions during early infection stages (Ahuja et al., 2012). However, the mechanisms employed by camalexin to exert its toxicity remain unknown.

1.2.6 Implication of long noncoding RNAs in plant defenses

Genomic and transcriptomic data available from higher eukaryotes including humans and diverse plant species have revealed that genes encoding proteins only constitute a small fraction, about 1.5%, of the total genetic material (Mignone et al., 2002). Indeed, it is now known that most portions of higher eukaryotic genomic DNA function in modulation of gene expression, exerted at either the level of transcription, controlling whether or not a gene is transcribed and the extent of transcription, or post-transcriptionally, regulating the fate of transcripts, including their stability, translation efficiency and subcellular localization (Mignone et al., 2002). Increasingly, recent evidence now indicates that significant proportions of the pervasively transcribed unannotated genomic regions are noncoding RNAs (ncRNAs) performing important roles in diverse biological processes, including response to abiotic and biotic stresses (Au et al., 2011, Wang and Chang, 2011, Wilusz et al., 2009). Based on their length, ncRNAs can be divided into small ncRNAs, generally less than 200 base pairs (bp), this class includes micro RNAs (miRNAs), short interfering RNAs (siRNAs), small nucleolar RNAs (snoRNAs) and small nuclear RNAs (snRNAs); and long ncRNAs (lncRNAs), which comprise longer molecules above 200 bp in length. lncRNAs are less characterized and appear less conserved than mRNA genes.

Depending on their orientation and genomic location with respect to neighboring protein-coding genes, lncRNAs are classified into natural antisense ncRNAs, intergenic, intronic and adjacent-lncRNAs (associated with promoter and terminator regions). lncRNAs possess signatures archetypal of mRNA molecules including presence of a 5`-cap, polyadenylation, and splicing, however, with few or no ORFs (Au et al., 2011). Plant lncRNAs are mainly transcribed by RNA polymerase II or III but in some cases polymerase IV/V may be involved. lncRNAs are potent gene transcription regulators, acting in *cis*- and *trans*-, and play regulatory roles during tissue development and plant responses to external stimuli (Kim and Sung, 2012). To date, short ncRNAs are the best characterized ncRNAs, whereas, research focusing on

lncRNAs is still in its infancy, especially in plants (Liu et al., 2015). Furthermore, studies of lncRNAs in plants are still lagging behind those of animals and humans (Zhu and Wang, 2012, Zhang and Chen, 2013, Bai et al., 2014). Nonetheless, next generation sequencing (NGS) technology, specifically, RNA sequencing (RNA-seq), has provided us with a complex and deeper perspective, and a complete picture of the RNA world, thereby making possible the resolution and discovery of lncRNA transcripts. Furthermore, application of strand-specific RNA-seq has made possible the detection of various transcripts including lncRNAs, providing strand information and transcription direction in both eukaryotic and prokaryotic genomes.

In contrast to short ncRNAs and mRNAs which are highly conserved, lncRNAs have poor conservation and their diverse regulatory mechanisms are still poorly understood (Wang and Chang, 2011). Even though the functions of only few plant lncRNAs have so far been characterized, numerous paradigms of their functionality are emerging (Wilusz et al., 2009). Most lncRNA transcripts identified in plants appear to be tissue-specific. In general, expression of lncRNAs is in response to diverse stimuli and they exhibit cell type-specific expression, suggesting that they are functional transcripts under considerable transcriptional control (Wang and Chang, 2011). Thus, as individual lncRNA molecules are transcribed at a particular time and specific location, they can function as molecular signals to interpret cellular context, integrate developmental cues or mediate responses to various abiotic or biotic stresses. Most lncRNAs serving as molecular signals possess regulatory functions. For instance, two plant lncRNA types, *COOLAIR* (COLD INDUCED LONG ANTISENSE INTRAGENIC RNA) and *COLDAIR* (COLD ASSISTED INTRONIC NONCODING RNA), have been identified in *Arabidopsis*, wherein, their transcription is triggered by winter cold resulting in epigenetic silencing of *FLOWERING LOCUS C* (*FLC*), a floral repressor (Kim et al., 2009a). The promoter of the antisense-lncRNA *COOLAIR* is cold inducible, and an increase in *COOLAIR* abundance leads to repression of *FLC* transcription and promotion of polycomb repressive complex 2 (PRC2 complex; polycomb) occupancy (Swiezewski et al., 2009). More recently, *COLDAIR* was shown to be essential in the vernalization-mediated epigenetic silencing of *FLC*, by directly recruiting PRC2 to *FLC* promoter regions (Heo and Sung, 2011). *COLDAIR*, is an intronic-lncRNA, transcribed from *FLC* and it directly interacts with *CURLY LEAF* (*CLF*), a component of polycomb. Collectively, *COOLAIR* and *COLDAIR* serve as transcriptional activity signals with spatial and temporal specificity (Liu et al., 2015).

Additionally, lncRNAs can act as precursors for miRNA biogenesis. In Arabidopsis, miRNAs such as miR869a and miR160c, are processed from *npc83* and *npc521* primary lncRNA sequences, respectively (Amor et al., 2009, Hirsch et al., 2006). Most of the precursor lncRNAs identified in (Amor et al., 2009) are responsive to different abiotic stresses including

salt stress, water stress and phosphate starvation. Another mechanism of lncRNA functioning involves lncRNAs acting as molecular decoys. In this archetype, transcribed lncRNAs bind and titrate away miRNAs from their authentic targets, such as mRNA transcripts, through partially perfect complementary interactions (Wu et al., 2013). These molecular decoys do not exert any additional functions, besides regulating miRNA functions, and are also referred to as competing endogenous RNAs (ceRNAs), which inhibit miRNA activity via target mimicry. Target mimicry was first discovered in plants but similar mechanisms or interactions were also identified in animal and human cells (Franco-Zorrilla et al., 2007, Liu et al., 2015). In Arabidopsis, *IPS1* (Induced by phosphate starvation1), an endogenous lncRNA, serves as a molecular decoy for miR399, blocking binding of mir399 to its intended targets (Franco-Zorrilla et al., 2007). mir399 is a phosphate starvation induced miRNA, and sequestration of mir399 by *IPS1* leads to increased accumulation of its authentic target, *PHO2* mRNA, leading to reduced phosphate content in the plant (Franco-Zorrilla et al., 2007). Additional functions of lncRNA involve acting as guides and scaffolds. LncRNAs acting either in *cis* (on neighboring genes) or *trans* (genes located distantly) can guide gene expression changes, however, the exact manner in which they carry out these functions is difficult to determine based on lncRNA sequences (Wang and Chang, 2011). In the scaffolds role, lncRNAs function as central platforms whereupon relevant molecular components are then assembled. Thus, various domains present within lncRNA sequences serve to bind distinct effectors in scaffolding complexes allowing precise control and specificity of intermolecular interactions in signaling processes (Wang and Chang, 2011).

In plants, lncRNAs involved in various biological processes have been identified, including fibre development in cotton, sexual reproductive process in rice, ripening-related in tomato, development regulation in Arabidopsis, and tissue-specific and development-related lncRNAs in maize (Wu et al., 2013, Li et al., 2014, Wang et al., 2015, Zhang et al., 2014, Zhu et al., 2015). Additionally, lncRNAs have been implicated in plant responses to abiotic external stimuli such as drought in *Populus trichocarpa*, salt stress in *Medicago truncatula*, and differentiation and stress responses in Arabidopsis (Amor et al., 2009, Shuai et al., 2014, Wen et al., 2007). Furthermore, lncRNAs are regulated in response to pathogen infection in plants. Recently, lncRNAs expressed in response to *Fusarium oxysporum* and Powdery mildew infection were reported in Arabidopsis and wheat, respectively (Xin et al., 2011, Zhu et al., 2014). Using RNA-inference knockdown lines, Zhu et al (Zhu et al., 2014) showed that five intergenic lncRNAs were related to disease resistance in Arabidopsis, following *F. oxysporum* infection. Arabidopsis lines in which the expression of these five lincRNAs was silenced exhibited either early disease symptoms or enhanced disease development after inoculation with *F. oxysporum* compared to control, uninoculated plants. Their study also suggested a

possible direct interaction between transcription factors and antisense-lncRNA promoter regions, underscoring the functionality of lncRNAs in plant responses during pathogen attack. Indeed, in addition to identifying lncRNA genomic locations in relation to protein-coding genes, determining the conditions under which they are expressed and within which plant cells or tissues they function is key in providing insights into their diversity and biological roles (Atkinson et al., 2012).

1.3 Soft Rot Enterobacteria

The soft rot enterobacteria group of phytopathogens (SRE) comprises broad host range pathogens that employ a necrotrophic mode of nutrition by producing PCWDEs which macerate host tissues allowing these pathogens to obtain nutrients from the dead or dying cells. The SRE are clustered in the *Enterobacteriaceae* family and they mainly belong to two genera, *Pectobacterium* and *Dickeya*. These genera represent necrotrophic phytopathogenic bacteria causing wilt, blackleg and rot diseases of plants (Davidsson et al., 2013, Charkowski et al., 2012). The taxonomical classification of these genera has been revised a number of times over the years, and members of both these genera were formerly classified under the genus *Erwinia* (Hauben et al., 1998, Naum et al., 2011, Gardan et al., 2003, Ma et al., 2007). SRE are found worldwide in diverse agricultural regions infecting various plants and have been isolated from different sources including, soil, water and plant hosts, but their main host is potato (*Solanum tuberosum*) and they cause tuber soft rot and blackleg diseases in potato stems resulting in significant losses (Pérombelon, 2002, Perombelon and Kelman, 1980, Charkowski et al., 2012).

Currently, the genus *Pectobacterium* can be subdivided into five different clades: *P. carotovorum* subsp. *carotovorum*, *P. atrosepticum*, *P. betavasculorum*, *P. wasabiae* and *P. carotovorum* subsp. *brasiliense* (Ma et al., 2007, Hauben et al., 1998, Nabhan et al., 2012, Gardan et al., 2003, Duarte et al., 2004). Within the genus *Pectobacterium*, *P. wasabiae*, *P. atrosepticum*, *P. carotovorum* subsp. *carotovorum* and *brasiliense* are responsible for causing wilt, rot and blackleg of stems, and tuber soft rot symptoms in potato (Ngadze et al., 2012, Kim et al., 2009b, Baghaee-Ravari et al., 2011, Pitman et al., 2010, van der Merwe et al., 2010). However, unlike *P. atrosepticum* which is more host-specific (mainly restricted to potato) and localized in temperate climates, *Pectobacterium* species such as members (subspecies) of *P. carotovorum*, tend to have a wide host range and are broadly distributed globally. For instance, *P. carotovorum* subsp *brasiliense* is emerging as a highly aggressive broad host range soft rot pathogen of global concern with the potential to cause severe losses in the potato industry. *P. carotovorum* subsp *brasiliense* has been isolated from potato plants or tubers grown in different climatic regions, including Canada, New Zealand, Netherlands,

Switzerland, Brazil, Israel, and in African countries including South Africa, Kenya and Zimbabwe (De Boer et al., 2012, Panda et al., 2012, Leite et al., 2014, Onkendi and Moleleki, 2014, Werra et al., 2015, Duarte et al., 2004, Ngadze et al., 2012).

The major common source for spread of soft rot and blackleg diseases are latently infected seed tubers, however, airborne sources such as dispersal of diseased plants by insects over long distances and surface water such as used for irrigation are additional sources for pathogen inoculum (Czajkowski et al., 2011). Soft rot pectobacteria are found on potato plant surfaces, usually in tuber lenticels or on roots, and or colonizing vascular tissue once inside the plant. Currently there are no available control measures of blackleg or soft rot diseases, once rotting symptoms appear there is nothing farmers can do to prevent disease progression in infected plants (Charkowski, 2015). However, use of copper sprays and good crop management during the growing season contribute to control of stem rot (Czajkowski et al., 2011, Charkowski, 2015).

1.3.1 *Pectobacterium* species virulence factors

Pectobacterium species are pectolytic Gram-negative, motile, facultative anaerobic phytopathogens belonging to the *γ-Proteobacteria* subdivision. Their primary virulence determinants are PCWDEs such as pectinases, cellulases, and proteases, secreted through the type II secretion system (T2SS), and are responsible for plant tissue maceration enabling *Pectobacterium* species to colonize and establish infection in various host plants. Among these exoenzymes, pectinases are the main PCWDEs involved in pathogenesis and disease establishment (Toth et al., 2003). Pectinases such as pectate lyases (Pels) break down and make use of pectins in plant cell walls and middle lamella of leaves leading to tissue collapse, cell damage and leakage (Pérombelon, 2002, Toth et al., 2003). Genes encoding PCWDEs are distributed among pectobacteria and are mainly clustered in their core genome (Glasner et al., 2008, Toth et al., 2006).

Favorable environmental conditions such as oxygen limitation, free water, high moisture and optimal temperature induce the initiation of soft rot disease as it triggers the switch of soft rot pectobacteria from a biotrophic asymptomatic state to a necrotrophic feeding mode (Pérombelon, 2002, Toth and Birch, 2005). Additionally, this switch in lifestyle is density dependent only occurring when the bacterial cell population is high. Furthermore, lifestyle switching is under the regulation of quorum sensing (QS), mediated by *expI* and *carI* genes through the production of acyl homoserine lactones (AHLs) (Toth and Birch, 2005, Toth et al., 2003). Thus, the synthesis of PCWDEs is tightly modulated by QS thus preventing premature elicitation of host defenses due to the action of PCWDEs before the critical bacterial

population density is attained, thereby, enhancing the pathogens successful infiltration (Davidsson et al., 2013). Furthermore, a key regulator of PCWDE production is KdgR, a transcriptional repressor, which interacts with cell wall breakdown products thereby accelerating synthesis of PCWDEs (Toth and Birch, 2005, Toth et al., 2003). Additional regulators of PCWDE synthesis include ExpRI, Crp, RexZ and H-NS, which appear conserved in *Pectobacterium* species (Glasner et al., 2008). Interestingly, *P. carotovorum* subsp *carotovorum* and *brasiliense* encode additional putative PCWDEs compared to other SRE such as *P. atrosepticum* (Glasner et al., 2008), which possibly contribute to the enhanced aggressiveness in *P. c. brasiliense* pathogenesis on potato plants and other plant species.

Other PCWDEs such as proteases are produced through a type 1 secretion system (T1SS), and they seem to contribute minimally to pathogenicity (Toth et al., 2006). In *Pectobacterium*, plant extracts and AHLs, lead to upregulation of the T1SS and the proteases may act in attacking proteins in plant cell walls or degradation of enzymes produced by the pathogen affecting their activity (Charkowski et al., 2012). With the exception of *P. wasabiae*, pectobacteria encode a type III secretion system (T3SS), but, the T3SS of pectobacteria seems not essential for growth on potato and establishment of a successful infection (Ma et al., 2007, Kim et al., 2009b). However, in *P. carotovorum* and *P. c. brasiliense*, the T3SS appears to contribute to virulence leading to elicitation of a T3SS-dependant HR cell death response in tobacco plants inoculated with these species (Glasner et al., 2008, Kim et al., 2011, Rantakari et al., 2001). The type IV secretion system (T4SS), required for plasmid conjugation is also present in some genomes of *Pectobacterium* including *P. c. brasiliense*, *P*. *atrosepticum* and some strains of *P. wasabiae*. Additional roles of the T4SS include translocation of DNA and/or proteins into host cells and secretion of proteins, in a similar manner to the T3SS, into the extracellular environment (Christie et al., 2005, Glasner et al., 2008). Additional secretion systems present in soft rot pathogens include the type V and VI secretion systems. The type V secretion system (T5SS) is the simplest among the other secretion systems, it includes auto-transporter and two-component secretion and it plays roles in pathogenesis of several bacterial pathogens (Henderson et al., 2004). In *Dickeya* spp. HecA hemagglutinin which promotes attachment to leaves, is secreted by the T5SS (Rojas et al., 2002). Finally the type VI secretion system (T6SS) which is common in Gram-negative bacteria is also encoded in soft rot pectobacteria, however, effector proteins secreted through the T6SS are yet to be discovered in pectobacteria (Davidsson et al., 2013). The T5SS and T6SS, act in adherence and have roles in competition with other microbes (Charkowski et al., 2012).

Additionally, important pathogenicity determinants in the SRE include flagella, adhesion and motility. Bacterial flagella have significant roles in motility and to an extent adhesion. Soft rot pectobacteria can swim or swarm, and they are motile during infection. Motility in SRE is via the flagella, and is essential for virulence in *Dickeya* and *Pectobacterium* species, and the flagella component is considered a sub-type of the T3SS (Charkowski et al., 2012). In addition, pectobacteria encode a wide array of methyl-accepting chemotaxis (MCP) receptors. Chemotaxis enables movement of bacterial cells toward certain stimuli through MCP sensing (Charkowski et al., 2012, Glasner et al., 2008, Lux and Shi, 2004). Thus, motility (including the contribution of flagella) and chemotaxis are important and essential in soft rot pectobacteria pathogenesis.

1.3.2 Survival strategies in nutrient limiting conditions

Currently, little is understood about survival strategies of soft rot pectobacteria between growing seasons, regardless of these pathogens having been studied for decades (Davidsson et al., 2013). Depending on soil temperature and other environmental conditions such as pH and moisture, survival of soft rots in soil is limited to weeks or a few months (up to six months). Thus, overwintering of pectobacteria in soil is considered unlikely (Czajkowski et al., 2011). However, decomposing plant material in soil can promote survival of pectobacteria (Czajkowski et al., 2011, Perombelon and Kelman, 1980). In the plant, some *Pectobacterium* species (such as *P. atrosepticum*) can infect roots and move through the vascular system into progeny tubers and cause infection. Alternatively, they move upwards colonizing stems and can either result in infection and disease or they survive latently within the nutrient limiting xylem vessels without necessarily causing blackleg disease.

The ability of SRE to contend with, survive or thrive in fluctuating complex and dynamic environments can be attributed to the pathogens ability to employ intracellular small molecules for signaling such as cyclic diguanylate (c-di-GMP), in sensing their physiological condition in order to adapt to the changes in their environment (Charkowski et al., 2012, Jenal and Malone, 2006). Consequently, c-di-GMP signaling network mediates bacterial adaptive responses. Furthermore, SRE bacterial cells can use intercellular signaling or communication by employing AHL-mediated quorum sensing (QS) to communicate with other bacterial cells (Jenal and Malone, 2006, Charkowski et al., 2012). In a cell density dependent manner, AHL-mediated QS controls the initiation of adaptive responses in bacteria required for persistence and ability to survive in unfavorable conditions. Bacterial survival under stress or starvation conditions requires drastic changes in gene expression necessary for stress response activation (adaptive response) (Gorshkov et al., 2010, Yildiz and Schoolnik, 1998). *P. atrosepticum* was shown to undergo a reversible transition into viable, but non-culturable

(VBN) dormant forms, allowing population maintenance and persistence under stress or starvation conditions (Gorshkov et al., 2009). For instance, during prolonged starvation, *P. atrosepticum* either significantly increases or decreases its population density (to ~10^6 CFU ml^{-1}) as it transforms into VBN forms, upon which AHL-dependent QS signaling is activated (Petrova et al., 2013, Gorshkov et al., 2010).

In Gram negative bacteria including soft rot pectobacteria, the adaptive (or stress) response is globally under the regulation of *rpoS*, which when activated expresses an alternative sigma factor during starvation (nutrient deprivation) conditions (Lange and Hengge-Aronis, 1991). The alternative sigma factor RpoS controls about 10% of *Escherichia coli* genes including genes related to cell morphology (cell envelope modification), stress resistance, virulence, metabolism and lysis, which prepare the bacterium for survival in stress-related conditions (Navarro Llorens et al., 2010, Yildiz and Schoolnik, 1998). Starvation conditions which trigger synthesis of RpoS include decreases in levels of carbon, phosphorus or nitrogen and amino acid starvation (Navarro Llorens et al., 2010). Thus, RpoS is the regulator of genes induced in stationary phase and is necessary for survival of bacteria during stress and starvation. Besides AHL-signaling, another important auto-inducer, in the bacterial adaptation process under harsh or stress conditions is ppGpp (alarmone guanosin 3`,5`-bidiphosphate) produced and regulated by RelA synthetase and/or SpoT synthetase/hydrolase (Petrova et al., 2013). Together RelA and SpoT sense carbon, iron, phosphorus, and fatty acid depravation and they trigger ppGpp synthesis when these nutrients are low. Thus, the alarmone ppGpp rapidly responds to various stresses leading to the induction of *rpoS* expression under nutrient deficient conditions (Hengge-Aronis, 2002).

Taken together, these adaptive response processes result in downregulation of rRNA biosynthesis, DNA replication, ribosomal proteins and upregulation of RpoS levels, amino acid biosynthesis and stress-related proteins in a phenomenon known as stringent response (Magnusson et al., 2005). The bacterial stringent response is reversed when bacterial cells encounter favorable environmental conditions (Navarro Llorens et al., 2010). Finally the stringent response plays significant roles in various bacterial processes including biofilm formation, quorum sensing and virulence regulation (Navarro Llorens et al., 2010). Interestingly, Petrova *et al.* (Petrova et al., 2013) showed that *P. atrosepticum* cells remain virulent during the stringent response, and this might have important implications in control strategies against soft rot pathogens in the potato industry.

1.3.2.1 Role of small RNAs in bacterial adaptive responses in nutrient deficient conditions

The process of bacterial adaptive response to stress conditions is coupled with induction of genes encoding essential regulatory elements required for realization of the stringent response. Among the necessary regulators are small RNAs (sRNAs) which have important functions in the regulation of numerous biological processes including acid resistance, translational quality control, and processing of tRNA molecules and other RNAs (Livny and Waldor, 2007, Waters and Storz, 2009). Importantly, most sRNAs act as post-transcriptional regulators interacting with various genes during stress response and/or recovery from stress (Gottesman et al., 2006). Most sRNAs have been identified computationally by means of predictive bioinformatics approaches, while numerous others have been detected or isolated through experimental analyses (Livny and Waldor, 2007). Bacterial sRNAs are relatively small, ranging in size between 50 and 500 nucleotides in length. Most sRNAs act via either extensive or limited base pairing with their target mRNAs controlling their translation and/or stability (Gottesman and Storz, 2011).

In prokaryotes, base pairing sRNAs can be grouped into two classes, antisense and intergenic sRNAs. Antisense (*cis*-encoded) sRNAs are transcribed from regions directly corresponding to their target genes on the opposite strand, thus, they are able to bind extensively with their mRNA targets forming complete complementarity of at least 75 nucleotides, and they have been studied in bacteria for many years and described to regulate plasmid and phage functions, and maintenance of mobile element copy numbers (Majdalani et al., 2005, Wagner et al., 2002). Furthermore, antisense sRNAs function in physiological roles including translational repression and/or promoting degradation of mRNAs that encode proteins which are toxic at elevated levels (Gerdes and Wagner, 2007). Other *cis*-encoded antisense sRNAs regulate the expression of genes within an operon on the complementary strand. For instance, *gadY* antisense sRNA in *E. coli* base-pairs with the 3`UTR regions of the *gadXW* mRNA during stationary phase, thereby stabilizing and enhancing transcription of *gadX* (Tramonti et al., 2008, Opdyke et al., 2004). Identification of antisense sRNAs especially those not associated with bacterial mobile elements is still far from complete, and more antisense sRNAs with different mechanisms of action are likely to continue being characterized in diverse bacterial species (Waters and Storz, 2009).

The other class comprises base pairing intergenic sRNAs which have limited complementarity with their target genes, usually around 10-25 nucleotides. These sRNAs, are *trans*-encoded, and are functionally analogous to miRNAs in eukaryotes (Gottesman, 2005). *Trans*-encoded sRNAs often regulate negatively their target mRNA, and their base pairing with their targets

usually results in repression of protein levels via mRNA degradation and/or translational inhibition (Gottesman, 2005). To date, most of the characterized *trans*-encoded sRNAs typically bind to the 5` untranslated regions (5`UTR) of mRNAs usually occluding the ribosome-binding site (Livny and Waldor, 2007). Reduction of protein levels is attributed to ribosome binding inhibition and the RNase E mediated degradation of the sRNA-mRNA duplex is thought to enhance the repression and making it irreversible (Morita et al., 2006). Since sRNAs form limited base pairs with their targets, they can bind multiple mRNAs. Several *trans*-encoded sRNAs require the RNA chaperone Hfq for their regulatory functionality (Gottesman, 2005). In addition, Hfq modulates sRNA levels, and when sRNAs are not base paired with mRNAs, Hfq protects these inactive sRNAs from degradation.

Recognition of intergenic sRNAs in recent years and their importance in bacterial gene expression regulation, especially genes involved in stress responses and bacterial pathogenesis, has led to increased interest in deciphering the regulatory functions of these molecules. This led to genome-wide searches and identification of these sRNAs in *E. coli* and other diverse bacterial organisms (Majdalani et al., 2005). Unlike, antisense sRNAs which are often expressed constitutively, *trans*-encoded sRNAs are mostly expressed under specific growth conditions, usually in response to changes in the environment. Some examples of these sRNAs in *E. coli*, include, *ryhB* (induced during low iron), and *sgrS* (activated by sugar-phosphate stress) (Vanderpool and Gottesman, 2004, Massé and Gottesman, 2002).

Other sRNAs, usually located within the 5`UTR, such as riboswitches, form part of the mRNA transcripts that they regulate, and they act in *cis* (Waters and Storz, 2009, Richards and Vanderpool, 2011). Riboswitches regulate gene expression by adopting different conformations in response to cues including changes in metabolite concentration, elevated temperatures, uncharged tRNAs, stalled ribosomes and small molecule ligands. Thus, riboswitches can be denoted as metabolite sensors, as they directly regulate genes involved in the utilization of the metabolite. Increasing number of various riboswitches are continually being identified in diverse bacterial species and riboswitches have aided in the clarification of physiological roles of some gene products (Waters and Storz, 2009).

Regulation of bacterial responses to stress conditions is complex. sRNAs play regulatory roles in the general stress response in bacteria, by enhancing or repressing the gene expression of key regulators allowing bacteria to adapt in nutrient-limiting environments. For example, the regulation of RpoS, the master regulator of the general stress response in Gram negative bacteria, such as *E. coli*, is complex, and its multi-leveled regulation is mostly meditaed by sRNAs post transcriptionally. In *E. coli*, seven sRNAs were shown to be part of the RpoS regulatory circuit in response to nutrient deprivation, osmotic shock, oxidative stress and

changes in temperature; expression of sRNAs *dsrA*, *rprA* and *arcZ* increases translation of RpoS (Majdalani et al., 2001, Mandin and Gottesman, 2010, Sledjeski et al., 1996), whereas *oxyS*, *cyaR*, *chiX* and *dicF* sRNAs downregulate *rpoS* expression (Mandin and Gottesman, 2010). The mode of action of *cyaR* and *chiX* is suggested to involve counteraction of *dsrA*, *rprA* and *arcZ* activation. Furthermore, these repressor sRNAs are thought to bind Hfq, preventing it from binding *rpoS* mRNA or the RpoS inducing sRNAs (Mandin and Gottesman, 2010). The three positive regulators of RpoS, act by directly base-pairing with *rpoS* mRNA in the upper stem region of its 5`UTR stem-loop structure, thus, allowing increased translation by interfering with *rpoS* mRNA stem structure (Mandin and Gottesman, 2010, Majdalani et al., 1998). Regulation of RpoS by multiple sRNA molecules allows bacterial cells to respond and adapt timeously to different stress signals. In addition, ability of sRNAs to regulate many mRNA targets can have important roles in co-regulation of different targets during bacterial stress response. For example, in *E.* coli, *dsrA*, induces RpoS translation, but downregulates the expression of H-NS (global regulator of gene expression), and *oxyS* represses both RpoS and FhlA (a transcriptional activator).

Furthermore, understanding conditions controlling changes in sRNA amounts in bacterial cells is key in deciphering the roles of sRNAs (Repoila et al., 2003). For instance, regulation of iron homeostasis in *E. coli*, includes the sRNA regulator, *ryhB* (Massé and Gottesman, 2002). *ryhB* is activated during iron starvation conditions and it acts by down-regulating non-essential genes expressing iron storage proteins and iron-requiring enzymes such as the *sdhCDAB* operon which encodes succinate dehydrogenase, thereby redirecting intracellular iron usage. Additionally, *ryhb* enhances siderophore biosynthesis, thus, increasing the ability of cells to scavenge for iron from the environment (Richards and Vanderpool, 2011). However, iron abundance, leads to repression of *ryhB* synthesis by Fur (a negative transcriptional regulator) (Massé and Gottesman, 2002, Richards and Vanderpool, 2011). Thus, the iron-dependent expression of *ryhB* allows cells to sense concentrations of iron in the medium and adjust cellular iron utilization (Repoila et al., 2003).

Together, sRNA abundance varies with changing environmental conditions, and sRNA expression and stability integrates varying environmental cues in sRNA-dependent regulation of various genes and regulators which forms part of the coordinated internal responses in the general stress response and adaptation in bacteria (Repoila et al., 2003, Richards and Vanderpool, 2011).

1.4 Transcriptome profiling

Transcriptome profiling with next-generation DNA sequencing (NGS) technology, particularly RNA sequencing (RNA-seq) has become a powerful tool of choice employing a sequence-based approach for analyzing levels of gene expression among multiple samples. RNA-seq provides information and a digital measure of the abundance and presence of transcripts from both novel and known genes (Mortazavi et al., 2008). In addition, RNA-seq profiles the transcriptome directly and it allows single-base resolution. Unlike microarrays, RNA-seq does not require prior information about the transcriptome being surveyed. In general, most transcriptomics strategies are based either on hybridization or sequence-based approaches. Since the mid-1990s, prior to the advent of RNA-seq, large-scale gene expression studies were largely hybridization-based and employed microarray approaches (Marioni et al., 2008, Ozsolak and Milos, 2011). However, the RNA-seq resolution power, efficiency and low cost advantages as a method for differential gene expression profiling has led many biologists to employ RNA-seq opposed to microarrays. Furthermore, inherent limitations of microarrays such as high levels of background noise due to cross-hybridization artefacts, and dye-based detection flaws, make it difficult for microarrays to provide full coverage of all possible gene-models including unknown genes in large eukaryotic genomes (Mortazavi et al., 2008).

Indeed, transcriptomics research was revolutionized by the introduction of NGS technologies about a decade ago which allowed RNA analysis through massive scale cDNA sequencing (Metzker, 2010, Ozsolak and Milos, 2011). Newer sequencing methods that were developed after the automated Sanger method are collectively referred to as NGS, and they employ a combination of strategies including template preparation, sequencing and imaging, and downstream (genome) assembly and annotation (Metzker, 2010). Additionally, the major benefit of NGS platforms is production of enormous volumes of data at a low cost. The major commercially available NGS platforms are from Roche/454, Life Technologies/Ion torrent PGM, Illumina/Solexa (including MiSeq and HiSeq sequencers), and recently developed third generation sequencers (Pacific Biosciences and Oxford Nanopore) (Metzker, 2010). The various features and technical advantages of the different NGS platforms in particular applications have been extensively reviewed (Metzker, 2010, Quail et al., 2012, Branton et al., 2008, Liu et al., 2012, Hodkinson and Grice, 2015). Currently, the Illumina platforms dominate the market, with the Illumina HiSeq sequencers at the forefront, both in terms of amount of reads generated (read output) and lowest reagent costs (Quail et al., 2012, Liu et al., 2012). For example, the Illumina HiSeq 2500 produces 4 billion paired-end reads (125 bases per each read) in a single run (Hodkinson and Grice, 2015).

RNA-seq based approaches have provided a progressively complete understanding of the qualitative and quantitative aspects of transcripts expressed in both bacterial and eukaryotic genomes (Ozsolak and Milos, 2011). Thus, advances in RNA-seq approaches enable comprehensive transcriptome studies involving reliable detection and identification of full sets of expressed transcript classes including mRNAs from annotated genes and novel transcripts, splicing isoforms, as well as small and long ncRNA. Applications of RNA-seq in transcriptomics vary amongst studies. For instance, the nucleotide resolution power of RNA-seq allows comprehensive mapping of transcription start sites (TSS) making it possible to fully define various RNA molecules and identify adjacent regulatory promoter regions. Additionally, TSS mapping using RNA-seq provides extensive information on TSS, untranslated regions of mRNA transcripts, and previously unknown sRNA genes in bacteria, thus improving genome annotations (Sharma and Vogel, 2014). For example, a wide range of bacterial sRNAs have been identified using TSS mapping by employing differential 5`-end RNA-seq (Dugar et al., 2013, Soutourina et al., 2013). Additionally, RNA-seq can be combined with deepCAGE for efficient mapping of TSS in mammalian cells (Valen et al., 2009). Furthermore, a modification in the standard libraries for RNA-seq can allow for strand-specific RNA-seq, providing strand information during sequencing. Typically, standard RNA-seq libraries do not preserve information about transcription direction (Levin et al., 2010).

Strand-specific RNA-seq has various applications in transcriptome profiling studies, including discovery of antisense transcripts such as long noncoding-NATs (natural antisense transcripts) with potential regulatory roles. In addition, stranded RNA-seq can be useful in determining the transcribed strand for novel genes or ncRNAs, demarcating the exact boundaries of adjacent genes on opposite strands, enable the resolution of correct expression levels overlapping coding or noncoding transcripts (Levin et al., 2010, Mills et al., 2013). The most widely used method for strand-specific RNA-seq, especially when using paired-end sequencing, is the incorporation of dUTP during second-strand cDNA synthesis followed by degradation of the strand without dUTP (Levin et al., 2010). Collectively, TSS mapping and strand-specific RNA-seq can be employed in gene expression profiling. Thus, the most common and popular use of RNA-seq is differential expression analysis of genes or profiling transcript abundance between samples, regardless of the RNA-seq libraries used.

In conclusion, RNA-seq largely aims to capture and profile the nature and extent of a transcriptome by profiling all expressed genes, determine gene expression changes in response to particular conditions or stimuli, identify co-expression and co-regulated genes in bacterial regulons (*trans*-regulated) or operons (*cis*-regulated) (Sharma and Vogel, 2014).

Finally, using RNA-seq structures of transcribed genes can be resolved, including their splice junctions and 5` and 3` regions, also RNA-seq can measure alternative splicing.

1.5 Study objectives

The overall aim of this study was to identify defense-related genes and long intergenic noncoding RNA transcripts (lincRNAs) in potato in response to *P. c. brasiliense* infection. To do this, strand-specific RNA-seq and time-course RNA-seq analyses were used to identify differentially expressed genes and lincRNAs between a susceptible potato cultivar (*S. tuberosum* cv Valor) and a tolerant cultivar (*S. tuberosum* cv BP1). The time-course study was carried out over 3 days and samples were obtained at 0, 6, 12, 24 and 72 hours post inoculation (hpi). Furthermore, the purpose of this study was to identify starvation-related small RNAs potentially important in *P. atrosepticum* adaptive responses.

The specific objectives were:

1. To determine the global potato defense transcriptome by identifying differentially expressed genes between a susceptible and a tolerant potato cultivar following infection with *P. c. brasiliense*.
2. To perform functional enrichment analysis (Gene ontology analysis) to characterize differentially expressed genes and determine key defense genes in potato against *P. c. brasiliense*.
3. To computationally identify and characterize candidate novel genes using potato RNA-seq transcriptome data.
4. To design and implement a bioinformatics pipeline for lincRNA identification using potato strand-specific RNA-seq data.
5. To determine the set of lincRNA transcripts responsive to *P. c. brasiliense* infection.
6. To design and implement a bioinformatics approach for identification of bacterial small RNAs using strand-specific RNA-seq data.
7. To determine the role of sRNAs in *Pectobacterium* adaptive responses

1.6 References

ABUQAMAR, S., CHAI, M.-F., LUO, H., SONG, F. & MENGISTE, T. 2008. Tomato protein kinase 1b mediates signaling of plant responses to necrotrophic fungi and insect herbivory. *The Plant Cell,* 20**,** 1964-1983.

AHUJA, I., KISSEN, R. & BONES, A. M. 2012. Phytoalexins in defense against pathogens. *Trends in Plant Science,* 17**,** 73-90.

AMOR, B. B., WIRTH, S., MERCHAN, F., LAPORTE, P., D'AUBENTON-CARAFA, Y., HIRSCH, J., MAIZEL, A., MALLORY, A., LUCAS, A. & DERAGON, J. M. 2009. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Research,* 19**,** 57-69.

ASAI, T., TENA, G., PLOTNIKOVA, J., WILLMANN, M. R., CHIU, W.-L., GOMEZ-GOMEZ, L., BOLLER, T., AUSUBEL, F. M. & SHEEN, J. 2002. MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature,* 415**,** 977-983.

ATKINSON, S. R., MARGUERAT, S. & BÄHLER, J. Exploring long non-coding RNAs through sequencing. *Seminars in Cell and Developmental Biology*, 2012. Elsevier, 200-205.

AU, P. C. K., ZHU, Q.-H., DENNIS, E. S. & WANG, M.-B. 2011. Long non-coding RNA-mediated mechanisms independent of the RNAi pathway in animals and plants. *RNA Biology,* 8**,** 404-414.

BAGHAEE-RAVARI, S., RAHIMIAN, H., SHAMS-BAKHSH, M., LOPEZ-SOLANILLA, E., ANTÚNEZ-LAMAS, M. & RODRÍGUEZ-PALENZUELA, P. 2011. Characterization of Pectobacterium species from Iran using biochemical and molecular methods. *European Journal of Plant Pathology,* 129**,** 413-425.

BAI, Y., DAI, X., HARRISON, A. P. & CHEN, M. 2014. RNA regulatory networks in animals and plants: a long noncoding RNA perspective. *Briefings in Functional Genomics***,** elu017.

BETHKE, G., PECHER, P., ESCHEN-LIPPOLD, L., TSUDA, K., KATAGIRI, F., GLAZEBROOK, J., SCHEEL, D. & LEE, J. 2012. Activation of the Arabidopsis mitogen-activated protein kinase MPK11 by the flagellin-derived elicitor peptide, flg22. *Molecular Plant-Microbe Interactions,* 25**,** 471-480.

BETHKE, G., UNTHAN, T., UHRIG, J. F., PÖSCHL, Y., GUST, A. A., SCHEEL, D. & LEE, J. 2009. Flg22 regulates the release of an ethylene response factor substrate from MAP kinase 6 in Arabidopsis via ethylene signaling. *Proceedings of the National Academy of Sciences,* 106**,** 8067-8072.

BIRCH, P. R., BRYAN, G., FENTON, B., GILROY, E. M., HEIN, I., JONES, J. T., PRASHAR, A., TAYLOR, M. A., TORRANCE, L. & TOTH, I. K. 2012. Crops that feed the world 8: Potato: are the trends of increased global production sustainable? *Food Security,* 4**,** 477-508.

BOLLER, T. & FELIX, G. 2009. A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology,* 60**,** 379-406.

BRANTON, D., DEAMER, D. W., MARZIALI, A., BAYLEY, H., BENNER, S. A., BUTLER, T., DI VENTRA, M., GARAJ, S., HIBBS, A. & HUANG, X. 2008. The potential and challenges of nanopore sequencing. *Nature Biotechnology,* 26**,** 1146-1153.

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,* 107**,** 9452-9457.

BURLINGAME, B., MOUILLÉ, B. & CHARRONDIÈRE, R. 2009. Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. *Journal of Food Composition and Analysis,* 22**,** 494-502.

CHARKOWSKI, A., BLANCO, C., CONDEMINE, G., EXPERT, D., FRANZA, T., HAYES, C., HUGOUVIEUX-COTTE-PATTAT, N., SOLANILLA, E. L., LOW, D. & MOLELEKI, L. 2012. The role of secretion systems and small molecules in soft-rot enterobacteriaceae pathogenicity. *Annual Review of Phytopathology,* 50**,** 425-449.

CHARKOWSKI, A. O. 2015. Biology and control of Pectobacterium in potato. *American Journal of Potato Research,* 92**,** 223-229.

CHISHOLM, S. T., COAKER, G., DAY, B. & STASKAWICZ, B. J. 2006. Host-microbe interactions: shaping the evolution of the plant immune response. *Cell,* 124**,** 803-814.

CHRISTIE, P. J., ATMAKURI, K., KRISHNAMOORTHY, V., JAKUBOWSKI, S. & CASCALES, E. 2005. Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annual Review of Microbiology,* 59.

COLL, N., EPPLE, P. & DANGL, J. 2011. Programmed cell death in the plant immune system. *Cell Death and Differentiation,* 18**,** 1247-1256.

CONSORTIUM, P.G.S. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475**,** 189-195.

CRISTINA, M. S., PETERSEN, M. & MUNDY, J. 2010. Mitogen-activated protein kinase signaling in plants. *Annual Review of Plant Biology,* 61**,** 621-649.

CZAJKOWSKI, R., PEROMBELON, M. C., VAN VEEN, J. A. & VAN DER WOLF, J. M. 2011. Control of blackleg and tuber soft rot of potato caused by Pectobacterium and Dickeya species: a review. *Plant Pathology,* 60**,** 999-1013.

DAVIDSSON, P. R., KARIOLA, T., NIEMI, O. & PALVA, E. T. 2013. Pathogenicity of and plant immunity to soft rot pectobacteria. *Frontiers in Plant Science,* 4.

DE BOER, S., LI, X. & WARD, L. 2012. Pectobacterium spp. associated with bacterial stem rot syndrome of potato in Canada. *Phytopathology,* 102**,** 937-947.

DEL POZO, O., PEDLEY, K. F. & MARTIN, G. B. 2004. MAPKKKα is a positive regulator of cell death associated with both plant immunity and disease. *The EMBO Journal,* 23**,** 3072-3082.

DIXON, R. A. 2001. Natural products and plant disease resistance. *Nature,* 411**,** 843-847.

DODDS, P. N. & RATHJEN, J. P. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics,* 11**,** 539-548.

DUARTE, V., DE BOER, S., WARD, L. & OLIVEIRA, A. 2004. Characterization of atypical Erwinia carotovora strains causing blackleg of potato in Brazil. *Journal of Applied Microbiology,* 96**,** 535-545.

DUGAR, G., HERBIG, A., FÖRSTNER, K. U., HEIDRICH, N., REINHARDT, R., NIESELT, K. & SHARMA, C. M. 2013. High-resolution transcriptome maps reveal strain-specific regulatory features of multiple Campylobacter jejuni isolates. *PLoS Genetics,* 9**,** e1003495.

EL OIRDI, M., EL RAHMAN, T. A., RIGANO, L., EL HADRAMI, A., RODRIGUEZ, M. C., DAAYF, F., VOJNOV, A. & BOUARAB, K. 2011. Botrytis cinerea manipulates the antagonistic effects between immune pathways to promote disease development in tomato. *The Plant Cell,* 23**,** 2405-2421.

ELMORE, J. M., LIN, Z.-J. D. & COAKER, G. 2011. Plant NB-LRR signaling: upstreams and downstreams. *Current Opinion in Plant Biology,* 14**,** 365-371.

FRANCO-ZORRILLA, J. M., VALLI, A., TODESCO, M., MATEOS, I., PUGA, M. I., RUBIO-SOMOZA, I., LEYVA, A., WEIGEL, D., GARCÍA, J. A. & PAZ-ARES, J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics,* 39**,** 1033-1037.

FREEMAN, B. & BEATTIE, G. 2008. An overview of plant defenses against pathogens and herbivores. The Plant Health Instructor. doi: 10.1094. PHI-I-2008-0226-01.

FRERIGMANN, H., GLAWISCHNIG, E. & GIGOLASHVILI, T. 2015. The role of MYB34, MYB51 and MYB122 in the regulation of camalexin biosynthesis in Arabidopsis. *Frontiers in Plant Science,* 6.

GARDAN, L., GOUY, C., CHRISTEN, R. & SAMSON, R. 2003. Elevation of three subspecies of Pectobacterium carotovorum to species level: Pectobacterium atrosepticum sp. nov., Pectobacterium betavasculorum sp. nov. and Pectobacterium wasabiae sp. nov. *International Journal of Systematic and Evolutionary Microbiology,* 53**,** 381-391.

GERDES, K. & WAGNER, E. G. H. 2007. RNA antitoxins. *Current Opinion in Microbiology,* 10**,** 117-124.

GLASNER, J., MARQUEZ-VILLAVICENCIO, M., KIM, H.-S., JAHN, C., MA, B., BIEHL, B., RISSMAN, A., MOLE, B., YI, X. & YANG, C.-H. 2008. Niche-specificity and the variable fraction of the Pectobacterium pan-genome. *Molecular Plant-Microbe Interactions,* 21**,** 1549-1560.

GLAZEBROOK, J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology,* 43**,** 205-227.

GÓMEZ-GÓMEZ, L. & BOLLER, T. 2000. FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Molecular Cell,* 5**,** 1003-1011.

GORSHKOV, V., PETROVA, O., GOGOLEVA, N. & GOGOLEV, Y. 2010. Cell-to-cell communication in the populations of enterobacterium Erwinia carotovora ssp. atroseptica SCRI1043 during adaptation to stress conditions. *FEMS Immunology and Medical Microbiology,* 59**,** 378-385.

GORSHKOV, V. Y., PETROVA, O., MUKHAMETSHINA, N., AGEEVA, M., MULYUKIN, A. & GOGOLEV, Y. V. 2009. Formation of “nonculturable” dormant forms of the phytopathogenic enterobacterium Erwinia carotovora. *Microbiology,* 78**,** 585-592.

GOTTESMAN, S. 2005. Micros for microbes: non-coding regulatory RNAs in bacteria. *TRENDS in Genetics,* 21**,** 399-404.

GOTTESMAN, S., MCCULLEN, C., GUILLIER, M., VANDERPOOL, C., MAJDALANI, N., BENHAMMOU, J., THOMPSON, K., FITZGERALD, P., SOWA, N. & FITZGERALD, D. Small RNA regulators and the bacterial response to stress. Cold Spring Harbor Symposia on Quantitative Biology, 2006. Cold Spring Harbor Laboratory Press, 1-11.

GOTTESMAN, S. & STORZ, G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology,* 3**,** a003798.

GREENBERG, J. T. & YAO, N. 2004. The role and regulation of programmed cell death in plant–pathogen interactions. *Cellular Microbiology,* 6**,** 201-211.

HAUBEN, L., MOORE, E. R., VAUTERIN, L., STEENACKERS, M., MERGAERT, J., VERDONCK, L. & SWINGS, J. 1998. Phylogenetic position of phytopathogens within the Enterobacteriaceae. *Systematic and Applied Microbiology,* 21**,** 384-397.

HENDERSON, I. R., NAVARRO-GARCIA, F., DESVAUX, M., FERNANDEZ, R. C. & ALA'ALDEEN, D. 2004. Type V protein secretion pathway: the autotransporter story. *Microbiology and Molecular Biology Reviews,* 68**,** 692-744.

HENGGE-ARONIS, R. 2002. Signal transduction and regulatory mechanisms involved in control of the σS (RpoS) subunit of RNA polymerase. *Microbiology and Molecular Biology Reviews,* 66**,** 373-395.

HENRY, E., YADETA, K. A. & COAKER, G. 2013. Recognition of bacterial plant pathogens: local, systemic and transgenerational immunity. *New Phytologist,* 199**,** 908-915.

HEO, J. B. & SUNG, S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science,* 331**,** 76-79.

HERNÁNDEZ-BLANCO, C., FENG, D. X., HU, J., SÁNCHEZ-VALLET, A., DESLANDES, L., LLORENTE, F., BERROCAL-LOBO, M., KELLER, H., BARLET, X. & SÁNCHEZ-RODRÍGUEZ, C. 2007. Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *The Plant Cell,* 19**,** 890-903.

HIRSCH, J., LEFORT, V., VANKERSSCHAVER, M., BOUALEM, A., LUCAS, A., THERMES, C., D'AUBENTON-CARAFA, Y. & CRESPI, M. 2006. Characterization of 43 non-protein-coding mRNA genes in Arabidopsis, including the MIR162a-derived transcripts. *Plant Physiology,* 140**,** 1192-1204.

HODKINSON, B. P. & GRICE, E. A. 2015. Next-generation sequencing: a review of technologies and tools for wound microbiome research. *Advances in Wound Care,* 4**,** 50-58.

HUFFAKER, A., PEARCE, G. & RYAN, C. A. 2006. An endogenous peptide signal in Arabidopsis activates components of the innate immune response. *Proceedings of the National Academy of Sciences,* 103**,** 10098-10103.

HUFFAKER, A. & RYAN, C. A. 2007. Endogenous peptide defense signals in Arabidopsis differentially amplify signaling for the innate immune response. *Proceedings of the National Academy of Sciences,* 104**,** 10732-10736.

ICHIMURA, K., SHINOZAKI, K., TENA, G., SHEEN, J., HENRY, Y., CHAMPION, A., KREIS, M., ZHANG, S., HIRT, H. & WILSON, C. 2002. Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in Plant Science,* 7**,** 301-308.

ISHIHAMA, N., YAMADA, R., YOSHIOKA, M., KATOU, S. & YOSHIOKA, H. 2011. Phosphorylation of the Nicotiana benthamiana WRKY8 transcription factor by MAPK functions in the defense response. *The Plant Cell,* 23**,** 1153-1170.

JENAL, U. & MALONE, J. 2006. Mechanisms of cyclic-di-GMP signaling in bacteria. *Annual Reviews Genetics,* 40**,** 385-407.

JONES, J. D. & DANGL, J. L. 2006. The plant immune system. *Nature,* 444**,** 323-329.

KIM, D.-H., DOYLE, M. R., SUNG, S. & AMASINO, R. M. 2009a. Vernalization: winter and the timing of flowering in plants. *Annual Review of Cell and Developmental,* 25**,** 277-299.

KIM, E.-D. & SUNG, S. 2012. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends in Plant Science,* 17**,** 16-21.

KIM, H.-S., MA, B., PERNA, N. T. & CHARKOWSKI, A. O. 2009b. Phylogeny and virulence of naturally occurring type III secretion system-deficient Pectobacterium strains. *Applied and Environmental Microbiology,* 75**,** 4539-4549.

KIM, H.-S., THAMMARAT, P., LOMMEL, S. A., HOGAN, C. S. & CHARKOWSKI, A. O. 2011. Pectobacterium carotovorum elicits plant cell death with DspE/F but the P. carotovorum DspE does not suppress callose or induce expression of plant genes early in plant-microbe interactions. *Molecular Plant-Microbe Interactions,* 24**,** 773-786.

KO, J.-H., JEON, H.-W., KIM, W.-C., KIM, J.-Y. & HAN, K.-H. 2014. The MYB46/MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis. *Annals of Botany***,** mcu126.

KUNZE, G., ZIPFEL, C., ROBATZEK, S., NIEHAUS, K., BOLLER, T. & FELIX, G. 2004. The N terminus of bacterial elongation factor Tu elicits innate immunity in Arabidopsis plants. *The Plant Cell,* 16**,** 3496-3507.

LAI, Z. & MENGISTE, T. 2013. Genetic and cellular mechanisms regulating plant responses to necrotrophic pathogens. *Current Opinion in Plant Biology,* 16**,** 505-512.

LANGE, R. & HENGGE-ARONIS, R. 1991. Identification of a central regulator of stationary-phase gene expression in Escherichia coli. *Molecular Microbiology,* 5**,** 49-59.

LEITE, L. N., DE HAAN, E., KRIJGER, M., KASTELEIN, P., VAN DER ZOUWEN, P., VAN DEN BOVENKAMP, G., TEBALDI, N. & VAN DER WOLF, J. 2014. First report of potato blackleg caused by Pectobacterium carotovorum subsp. brasiliensis in the Netherlands. *New Diseases Reports,* 29**,** 24.

LEVIN, J. Z., YASSOUR, M., ADICONIS, X., NUSBAUM, C., THOMPSON, D. A., FRIEDMAN, N., GNIRKE, A. & REGEV, A. 2010. Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nature Methods,* 7**,** 709-715.

LI, L., EICHTEN, S. R., SHIMIZU, R., PETSCH, K., YEH, C.-T., WU, W., CHETTOOR, A. M., GIVAN, S. A., COLE, R. A. & FOWLER, J. E. 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology,* 15**,** R40.

LIU, L., LI, Y., LI, S., HU, N., HE, Y., PONG, R., LIN, D., LU, L. & LAW, M. 2012. Comparison of next-generation sequencing systems. *BioMed Research International,* 2012.

LIU, X., HAO, L., LI, D., ZHU, L. & HU, S. 2015. Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics, Proteomics and Bioinformatics*.

LIVNY, J. & WALDOR, M. K. 2007. Identification of small RNAs in diverse bacterial species. *Current Opinion in Microbiology,* 10**,** 96-101.

LU, D., WU, S., GAO, X., ZHANG, Y., SHAN, L. & HE, P. 2010. A receptor-like cytoplasmic kinase, BIK1, associates with a flagellin receptor complex to initiate plant innate immunity. *Proceedings of the National Academy of Sciences,* 107**,** 496-501.

LUX, R. & SHI, W. 2004. Chemotaxis-guided movements in bacteria. *Critical Reviews in Oral Biology and Medicine,* 15**,** 207-220.

MA, B., HIBBING, M. E., KIM, H.-S., REEDY, R. M., YEDIDIA, I., BREUER, J., BREUER, J., GLASNER, J. D., PERNA, N. T. & KELMAN, A. 2007. Host range and molecular phylogenies of the soft rot enterobacterial genera Pectobacterium and Dickeya. *Phytopathology,* 97**,** 1150-1163.

MAGNUSSON, L. U., FAREWELL, A. & NYSTRÖM, T. 2005. ppGpp: a global regulator in Escherichia coli. *Trends in Microbiology,* 13**,** 236-242.

MAJDALANI, N., CHEN, S., MURROW, J., ST JOHN, K. & GOTTESMAN, S. 2001. Regulation of RpoS by a novel small RNA: the characterization of RprA. *Molecular Microbiology,* 39**,** 1382-1394.

MAJDALANI, N., CUNNING, C., SLEDJESKI, D., ELLIOTT, T. & GOTTESMAN, S. 1998. DsrA RNA regulates translation of RpoS message by an anti-antisense mechanism, independent of its action as an antisilencer of transcription. *Proceedings of the National Academy of Sciences,* 95**,** 12462-12467.

MAJDALANI, N., VANDERPOOL, C. K. & GOTTESMAN, S. 2005. Bacterial small RNA regulators. *Critical Reviews in Biochemistry and Molecular Biology,* 40**,** 93-113.

MALINOVSKY, F. G., FANGEL, J. U. & WILLATS, W. G. 2015. The role of the cell wall in plant immunity. *Plant Cell Wall in Pathogenesis, Parasitism and Symbiosis***,** 38.

MANDIN, P. & GOTTESMAN, S. 2010. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *The EMBO Journal,* 29**,** 3094-3107.

MAO, G., MENG, X., LIU, Y., ZHENG, Z., CHEN, Z. & ZHANG, S. 2011. Phosphorylation of a WRKY transcription factor by two pathogen-responsive MAPKs drives phytoalexin biosynthesis in Arabidopsis. *The Plant Cell,* 23**,** 1639-1653.

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.

MARQUEZ-VILLAVICENCIO, M. D. P., GROVES, R. L. & CHARKOWSKI, A. O. 2011. Soft rot disease severity is affected by potato physiology and Pectobacterium taxa. *Plant Disease,* 95**,** 232-241.

MASSÉ, E. & GOTTESMAN, S. 2002. A small RNA regulates the expression of genes involved in iron metabolism in Escherichia coli. *Proceedings of the National Academy of Sciences,* 99**,** 4620-4625.

MCCARTHY, R. L., ZHONG, R. & YE, Z.-H. 2009. MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant and Cell Physiology,* 50**,** 1950-1964.

MENG, X. & ZHANG, S. 2013. MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology,* 51**,** 245-266.

MENGISTE, T. 2012. Plant immunity to necrotrophs. *Annual Review of Phytopathology,* 50**,** 267-294.

MENGISTE, T., CHEN, X., SALMERON, J. & DIETRICH, R. 2003. The BOTRYTIS SUSCEPTIBLE1 gene encodes an R2R3MYB transcription factor protein that is required for biotic and abiotic stress responses in Arabidopsis. *The Plant Cell,* 15**,** 2551-2565.

METZKER, M. L. 2010. Applications of next generation sequencing sequencing technologies - the next generation. *Nature Reviews Genetics,* 11.

MEYERS, B. C., KOZIK, A., GRIEGO, A., KUANG, H. & MICHELMORE, R. W. 2003. Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. *The Plant Cell,* 15**,** 809-834.

MIEDES, E., VANHOLME, R., BOERJAN, W. & MOLINA, A. 2015. The role of the secondary cell wall in plant resistance to pathogens. *Plant Cell Wall in Pathogenesis, Parasitism and Symbiosis***,** 78.

MIGNONE, F., GISSI, C., LIUNI, S. & PESOLE, G. 2002. Untranslated regions of mRNAs. *Genome Biology,* 3**,** 1-10.

MILLS, J. D., KAWAHARA, Y. & JANITZ, M. 2013. Strand-specific RNA-Seq provides greater resolution of transcriptome profiling. *Current Genomics,* 14**,** 173.

MIYA, A., ALBERT, P., SHINYA, T., DESAKI, Y., ICHIMURA, K., SHIRASU, K., NARUSAKA, Y., KAWAKAMI, N., KAKU, H. & SHIBUYA, N. 2007. CERK1, a LysM receptor kinase, is essential for chitin elicitor signaling in Arabidopsis. *Proceedings of the National Academy of Sciences,* 104**,** 19613-19618.

MORITA, T., MOCHIZUKI, Y. & AIBA, H. 2006. Translational repression is sufficient for gene silencing by bacterial small noncoding RNAs in the absence of mRNA destruction. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 4858-4863.

MORTAZAVI, A., WILLIAMS, B. A., MCCUE, K., SCHAEFFER, L. & WOLD, B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods,* 5.

NABHAN, S., BOER, S., MAISS, E. & WYDRA, K. 2012. Taxonomic relatedness between Pectobacterium carotovorum subsp. carotovorum, Pectobacterium carotovorum subsp. odoriferum and Pectobacterium carotovorum subsp. brasiliense subsp. nov. *Journal of Applied Microbiology,* 113**,** 904-913.

NAUM, M., BROWN, E. W. & MASON-GAMER, R. J. 2011. Is a robust phylogeny of the enterobacterial plant pathogens attainable? *Cladistics,* 27**,** 80-93.

NAVARRO LLORENS, J. M., TORMO, A. & MARTÍNEZ-GARCÍA, E. 2010. Stationary phase in gram-negative bacteria. *FEMS Microbiology Reviews,* 34**,** 476-495.

NGADZE, E., BRADY, C. L., COUTINHO, T. A. & VAN DER WAALS, J. E. 2012. Pectinolytic bacteria associated with potato soft rot and blackleg in South Africa and Zimbabwe. *European Journal of Plant Pathology,* 134**,** 533-549.

OH, C.-S. & MARTIN, G. B. 2011. Effector-triggered immunity mediated by the Pto kinase. *Trends in Plant Science,* 16**,** 132-140.

ONKENDI, E. M. & MOLELEKI, L. N. 2014. Characterization of Pectobacterium carotovorum subsp. carotovorum and brasiliense from diseased potatoes in Kenya. *European Journal of Plant Pathology,* 139**,** 557-566.

OPDYKE, J. A., KANG, J.-G. & STORZ, G. 2004. GadY, a small-RNA regulator of acid response genes in Escherichia coli. *Journal of Bacteriology,* 186**,** 6698-6705.

OZSOLAK, F. & MILOS, P. M. 2011. RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics,* 12**,** 87-98.

PANDA, P., FIERS, M., ARMSTRONG, K. & PITMAN, A. 2012. First report of blackleg and soft rot of potato caused by Pectobacterium carotovorum subsp. brasiliensis in New Zealand. *New Disease Reports,* 26**,** 15.

PEDLEY, K. F. & MARTIN, G. B. 2005. Role of mitogen-activated protein kinases in plant immunity. *Current Opinion in Plant Biology,* 8**,** 541-547.

PÉROMBELON, M. 2002. Potato diseases caused by soft rot erwinias: an overview of pathogenesis. *Plant Pathology,* 51**,** 1-12.

PEROMBELON, M. C. & KELMAN, A. 1980. Ecology of the soft rot erwinias. *Annual Review of Phytopathology,* 18**,** 361-387.

PETROVA, O., GORSHKOV, V., DAMINOVA, A., AGEEVA, M., MOLELEKI, L. & GOGOLEV, Y. 2013. Stress response in Pectobacterium atrosepticum SCRI1043 under starvation conditions: adaptive reactions at a low population density. *Research in Microbiology,* 165**,** 119-127.

PITMAN, A. R., HARROW, S. A. & VISNOVSKY, S. B. 2010. Genetic characterisation of Pectobacterium wasabiae causing soft rot disease of potato in New Zealand. *European Journal of Plant Pathology,* 126**,** 423-435.

PITZSCHKE, A., SCHIKORA, A. & HIRT, H. 2009. MAPK cascade signalling networks in plant defence. *Current Opinion in Plant Biology,* 12**,** 421-426.

QUAIL, M. A., SMITH, M., COUPLAND, P., OTTO, T. D., HARRIS, S. R., CONNOR, T. R., BERTONI, A., SWERDLOW, H. P. & GU, Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics,* 13**,** 1.

RAMÍREZ, V., AGORIO, A., COEGO, A., GARCÍA-ANDRADE, J., HERNÁNDEZ, M. J., BALAGUER, B., OUWERKERK, P. B., ZARRA, I. & VERA, P. 2011. MYB46

modulates disease susceptibility to Botrytis cinerea in Arabidopsis. *Plant Physiology,* 155**,** 1920-1935.

RANTAKARI, A., VIRTAHARJU, O., VÄHÄMIKO, S., TAIRA, S., PALVA, E., SAARILAHTI, H. & ROMANTSCHUK, M. 2001. Type III secretion contributes to the pathogenesis of the soft-rot pathogen Erwinia carotovora: partial characterization of the hrp gene cluster. *Molecular Plant-Microbe Interactions,* 14**,** 962-968.

REPOILA, F., MAJDALANI, N. & GOTTESMAN, S. 2003. Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: the RpoS paradigm. *Molecular Microbiology,* 48**,** 855-861.

RICHARDS, G. R. & VANDERPOOL, C. K. 2011. Molecular call and response: the physiology of bacterial small RNAs. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1809**,** 525-531.

ROBATZEK, S., BITTEL, P., CHINCHILLA, D., KÖCHNER, P., FELIX, G., SHIU, S.-H. & BOLLER, T. 2007. Molecular identification and characterization of the tomato flagellin receptor LeFLS2, an orthologue of Arabidopsis FLS2 exhibiting characteristically different perception specificities. *Plant Molecular Biology,* 64**,** 539-547.

ROJAS, C. M., HAM, J. H., DENG, W.-L., DOYLE, J. J. & COLLMER, A. 2002. HecA, a member of a class of adhesins produced by diverse pathogenic bacteria, contributes to the attachment, aggregation, epidermal cell killing, and virulence phenotypes of Erwinia chrysanthemi EC16 on Nicotiana clevelandii seedlings. *Proceedings of the National Academy of Sciences,* 99**,** 13142-13147.

SHAN, L., HE, P., LI, J., HEESE, A., PECK, S. C., NÜRNBERGER, T., MARTIN, G. B. & SHEEN, J. 2008. Bacterial effectors target the common signaling partner BAK1 to disrupt multiple MAMP receptor-signaling complexes and impede plant immunity. *Cell host and Microbe,* 4**,** 17-27.

SHAN, X., ZHANG, Y., PENG, W., WANG, Z. & XIE, D. 2009. Molecular mechanism for jasmonate-induction of anthocyanin accumulation in Arabidopsis. *Journal of Experimental Botany,* 60**,** 3849-3860.

SHARMA, C. M. & VOGEL, J. 2014. Differential RNA-seq: the approach behind and the biological insight gained. *Current Opinion in Microbiology,* 19**,** 97-105.

SHUAI, P., LIANG, D., TANG, S., ZHANG, Z., YE, C.-Y., SU, Y., XIA, X. & YIN, W. 2014. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa. *Journal of Experimental Botany***,** eru256.

SLEDJESKI, D., GUPTA, A. & GOTTESMAN, S. 1996. The small RNA, DsrA, is essential for the low temperature expression of RpoS during exponential growth in Escherichia coli. *The EMBO Journal,* 15**,** 3993.

SOUTOURINA, O. A., MONOT, M., BOUDRY, P., SAUJET, L., PICHON, C., SISMEIRO, O., SEMENOVA, E., SEVERINOV, K., LE BOUGUENEC, C. & COPPÉE, J.-Y. 2013. Genome-Wide Identification of Regulatory RNAs in the Human Pathogen Clostridium difficile. *PLoS Genetics,* 9**,** e1003493.

SUAREZ-RODRIGUEZ, M. C., ADAMS-PHILLIPS, L., LIU, Y., WANG, H., SU, S.-H., JESTER, P. J., ZHANG, S., BENT, A. F. & KRYSAN, P. J. 2007. MEKK1 is required for flg22-induced MPK4 activation in Arabidopsis plants. *Plant Physiology,* 143**,** 661-669.

SWIEZEWSKI, S., LIU, F., MAGUSIN, A. & DEAN, C. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature,* 462**,** 799-802.

TAO, Y., XIE, Z., CHEN, W., GLAZEBROOK, J., CHANG, H.-S., HAN, B., ZHU, T., ZOU, G. & KATAGIRI, F. 2003. Quantitative nature of Arabidopsis responses during compatible and incompatible interactions with the bacterial pathogen Pseudomonas syringae. *The Plant Cell,* 15**,** 317-330.

TENA, G., BOUDSOCQ, M. & SHEEN, J. 2011. Protein kinase signaling networks in plant innate immunity. *Current Opinion in Plant Biology,* 14**,** 519-529.

TOTH, I. K., BELL, K. S., HOLEVA, M. C. & BIRCH, P. R. 2003. Soft rot erwiniae: from genes to genomes. *Molecular Plant Pathology,* 4**,** 17-30.

TOTH, I. K. & BIRCH, P. R. 2005. Rotting softly and stealthily. *Current Opinion in Plant Biology,* 8**,** 424-429.

TOTH, I. K., PRITCHARD, L. & BIRCH, P. R. 2006. Comparative genomics reveals what makes an enterobacterial plant pathogen. *Annual Review of Phytopathology,* 44**,** 305-336.

TRAMONTI, A., DE CANIO, M. & DE BIASE, D. 2008. GadX/GadW-dependent regulation of the Escherichia coli acid fitness island: transcriptional control at the gadY–gadW divergent promoters and identification of four novel 42 bp GadX/GadW-specific binding sites. *Molecular Microbiology,* 70**,** 965-982.

TSUDA, K., SATO, M., STODDARD, T., GLAZEBROOK, J. & KATAGIRI, F. 2009. Network properties of robust immunity in plants. *PLoS Genetics,* 5**,** e1000772.

VALEN, E., PASCARELLA, G., CHALK, A., MAEDA, N., KOJIMA, M., KAWAZU, C., MURATA, M., NISHIYORI, H., LAZAREVIC, D. & MOTTI, D. 2009. Genome-wide detection and analysis of hippocampus core promoters using DeepCAGE. *Genome Research,* 19**,** 255-265.

VAN DER MERWE, J. J., COUTINHO, T. A., KORSTEN, L. & VAN DER WAALS, J. E. 2010. Pectobacterium carotovorum subsp. brasiliensis causing blackleg on potatoes in South Africa. *European Journal of Plant Pathology,* 126**,** 175-185.

VANDERPOOL, C. K. & GOTTESMAN, S. 2004. Involvement of a novel transcriptional activator and small RNA in post-transcriptional regulation of the glucose phosphoenolpyruvate phosphotransferase system. *Molecular Microbiology,* 54**,** 1076-1089.

VANETTEN, H. D., MANSFIELD, J. W., BAILEY, J. A. & FARMER, E. E. 1994. Two classes of plant antibiotics: Phytoalexins versus" Phytoanticipins". *The Plant Cell,* 6**,** 1191.

VERONESE, P., NAKAGAMI, H., BLUHM, B., ABUQAMAR, S., CHEN, X., SALMERON, J., DIETRICH, R. A., HIRT, H. & MENGISTE, T. 2006. The membrane-anchored BOTRYTIS-INDUCED KINASE1 plays distinct roles in Arabidopsis resistance to necrotrophic and biotrophic pathogens. *The Plant Cell,* 18**,** 257-273.

WAGNER, E. G. H., ALTUVIA, S. & ROMBY, P. 2002. 12 Antisense RNAs in bacteria and their genetic elements. *Advances in Genetics,* 46**,** 361-398.

WANG, K. C. & CHANG, H. Y. 2011. Molecular mechanisms of long noncoding RNAs. *Molecular Cell,* 43**,** 904-914.

WANG, M., YUAN, D., TU, L., GAO, W., HE, Y., HU, H., WANG, P., LIU, N., LINDSEY, K. & ZHANG, X. 2015. Long noncoding RNAs and their proposed functions in fibre development of cotton (Gossypium spp.). *New Phytologist*.

WANG, X., JIANG, N., LIU, J., LIU, W. & WANG, G.-L. 2014. The role of effectors and host immunity in plant–necrotrophic fungal interactions. *Virulence,* 5**,** 722-732.

WATERS, L. S. & STORZ, G. 2009. Regulatory RNAs in bacteria. *Cell,* 136**,** 615-628.

WEN, J., PARKER, B. J. & WEILLER, G. F. 2007. In silico identification and characterization of mRNA-like noncoding transcripts in Medicago truncatula. *In silico Biology,* 7**,** 485-505.

WERRA, P. D., BUSSEREAU, F., KEISER, A. & ZIEGLER, D. 2015. First report of potato blackleg caused by Pectobacterium carotovorum subsp. brasiliense in Switzerland. *Plant Disease,* 99**,** 551-552.

WILUSZ, J. E., SUNWOO, H. & SPECTOR, D. L. 2009. Long noncoding RNAs: functional surprises from the RNA world. *Genes and Development,* 23**,** 1494-1504.

WU, H.-J., WANG, Z.-M., WANG, M. & WANG, X.-J. 2013. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiology,* 161**,** 1875-1884.

XIN, M., WANG, Y., YAO, Y., SONG, N., HU, Z., QIN, D., XIE, C., PENG, H., NI, Z. & SUN, Q. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biology,* 11**,** 61.

YAMAGUCHI, Y., PEARCE, G. & RYAN, C. A. 2006. The cell surface leucine-rich repeat receptor for AtPep1, an endogenous peptide elicitor in Arabidopsis, is functional in

transgenic tobacco cells. *Proceedings of the National Academy of Sciences,* 103**,** 10104-10109.

YILDIZ, F. H. & SCHOOLNIK, G. K. 1998. Role of rpoS in stress survival and virulence of Vibrio cholerae. *Journal of Bacteriology,* 180**,** 773-784.

ZHANG, X., DAI, Y., XIONG, Y., DEFRAIA, C., LI, J., DONG, X. & MOU, Z. 2007. Overexpression of Arabidopsis MAP kinase kinase 7 leads to activation of plant basal and systemic acquired resistance. *The Plant Journal,* 52**,** 1066-1079.

ZHANG, Y.-C. & CHEN, Y.-Q. 2013. Long noncoding RNAs: New regulators in plant development. *Biochemical and Biophysical Research Communications,* 436**,** 111-114.

ZHANG, Y.-C., LIAO, J.-Y., LI, Z.-Y., YU, Y., ZHANG, J.-P., LI, Q.-F., QU, L.-H., SHU, W.-S. & CHEN, Y.-Q. 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biology,* 15**,** 512.

ZHU, B., YANG, Y., LI, R., FU, D., WEN, L., LUO, Y. & ZHU, H. 2015. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *Journal of Experimental Botany***,** erv203.

ZHU, Q.-H. & WANG, M.-B. 2012. Molecular functions of long non-coding RNAs in plants. *Genes,* 3**,** 176-190.

ZHU, Q. H., STEPHEN, S., TAYLOR, J., HELLIWELL, C. A. & WANG, M. B. 2014. Long noncoding RNAs responsive to Fusarium oxysporum infection in Arabidopsis. *New Phytologist,* 201**,** 574-584.

ZIPFEL, C. 2008. Pattern-recognition receptors in plant innate immunity. *Current Opinion in Immunology,* 20**,** 10-16.

ZIPFEL, C., KUNZE, G., CHINCHILLA, D., CANIARD, A., JONES, J. D., BOLLER, T. & FELIX, G. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell,* 125**,** 749-760.

CHAPTER TWO

RNA-seq profiling reveals defense responses in a tolerant potato cultivar to stem infection by *Pectobacterium carotovorum* ssp. *brasiliense*

Stanford Kwenda[1], Tshepiso V. Motlolometsi[1], Paul R. J. Birch[2], Lucy N. Moleleki[1]

[1]Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0028, South Africa.

[2]The Division of Plant Sciences, College of Life Sciences, University of Dundee (The James Hutton Institute), Dundee DD25DA, Scotland, UK

This chapter has been prepared in the format of a manuscript and has been accepted for publication in the peer-reviewed journal Frontiers in Plant Science, and can be accessed under DOI: 10.3389/fpls.2016.01905. I performed the study design and conducted the experimental work, bioinformatics analyses, and drafted the manuscript. Tshepiso V. Motlolometsi assisted with RT-PCR validation of candidate novel CDS transcripts. Prof Lucy N. Moleleki conceived the study and was involved in the experimental design, helped to draft the manuscript, and she obtained funding which supported this work. Prof Paul R. J. Birch provided critical technical evaluation of the manuscript.

2.1 Abstract

Background: *Pectobacterium carotovorum* subsp. *brasiliense* is a member of the soft rot Enterobacteriaceae (SRE) family that causes tuber soft rot and blackleg diseases of stems in potato plants. Currently, there are no effective chemical strategies for the control of members of the SRE. Thus, an understanding of the inducible defense responses in stems of potato plants is important, particularly during colonization of the vascular system.

Results: Here, time-course RNA-sequencing analysis was used to compare expressed genes between a susceptible potato cultivar (*Solanum tubersoum* cv Valor) and a tolerant cultivar (*S. tuberosum* cv BP1) at 0, 6, 12, 24, and 72 h post-inoculation with *P. c. brasiliense*. In total, we identified 6,139 and 8,214 differentially expressed genes (DEGs) in the tolerant and susceptible cultivars, compared to mock-inoculated controls, respectively. Key DEGs distinguishing between tolerance and susceptibility were associated with negative regulation of cell death and plant-type cell wall organization/biogenesis biological processes in the tolerant and susceptible cultivars, respectively. Among these were DEGs involved in signaling (mainly MAPK cascade and ethylene pathway), defense-related transcription regulation including WRKY transcription factors, and downstream secondary cell biosynthesis.

Conclusion: Together, our results suggest that *S. tuberosum* cv BP1 likely employs quantitative defense response against *P.c brasiliense*. Overall, our study provides the first insight into the molecular basis of tolerance and/or resistance of potato stems to *P. c. brasiliense* infection.

2.2 Introduction

Potato ranks fourth, after rice (*Oryza sativa*), wheat (*Triticum* aestivum), and maize (*Zea* mays), as the most important human food crop worldwide (http://faostat.fao.org/site/339/default.aspx). However, cultivated potatoes, like many other plants, are exposed to diverse abiotic and biotic stresses. Some of the most important bacterial pathogens of potatoes belong to the soft rot enterobacteriaceae (SRE) consisting of *Dickeya* and *Pectobacterium* spp. In South Africa, *Pectobacterium carotovorum* subsp. *brasiliense* is the most widespread and aggressive soft rot enterobacterium, causing stem rot and blackleg in the field as well as tuber soft rot during post-harvest storage (van der Merwe et al., 2010). Incidentally, the global significance of *P. c. brasiliense* is growing with reports in countries such as Brazil, Canada, USA, New Zealand, China, and South Africa (Duarte et al., 2004;Glasner et al., 2008;van der Merwe et al., 2010;De Boer et al., 2012;Panda et al., 2012). Amongst the SRE are broad-host-range necrotrophic bacterial pathogens that employ plant cell wall degrading enzymes (PCWDEs) to macerate host tissues and obtain nutrients from dead cells (Davidsson et al., 2013). However, evidence suggests that soft rot bacteria can also exist as hemibiotrophs, living within the plant tissue (or in the surrounding environment) in an asymptomatic biotrophic state and only switching to a necrotrophic feeding mode when environmental conditions are favorable (Toth and Birch, 2005;Davidsson et al., 2013). The fact that the SRE localize deep inside the xylem or tuber lenticels makes effective control very difficult. Consequently, as with other vascular-dwelling pathogens, there are no efficient chemical control measures against SRE. Thus, the use of resistant cultivars remains the most desirable option of combating SRE (Charkowski, 2015).

Global gene expression studies of potato responses to environmental stresses such as drought, heat, and salinity (Massa et al., 2013;Gong et al., 2014) and biotic stresses as a result of fungal infections, predominantly caused by *Phytophthora infestans*, the causal agent of potato late blight, have been studied (Gyetvai et al., 2012;Gao et al., 2013;Massa et al., 2013).

Currently, little is known about the molecular basis of potato resistance to soft rot phytopathogens, with only a few commercial cultivars exhibiting tolerance to challenge by these bacteria (Charkowski, 2015). Fortunately, the availability of the potato genome sequence and next generation sequencing approaches such as RNA-seq (Wang et al., 2009;Consortium, 2011), now make it possible to conduct in-depth transcriptome studies in deciphering the potato defense transcriptome in response to soft rot bacterial infection, particularly against the emerging phytopathogen, *P. c. brasiliense*.

Invasion of plants by microbes activates plant immune responses which limit proliferation of pathogens and arrest disease establishment. Plant immune responses are complex and vary depending on whether the invading pathogen is of biotrophic or necrotrophic lifestyle (Mengiste, 2012). Thus, plant immune responses are composed of pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) pathways (Jones and Dangl, 2006). Generally PTI confers quantitative resistance in recognition of PAMPs (such as bacterial flagellin) and damage-associated molecular patterns (DAMPs) which mainly encompass degradation products from host cells due to the action of cell wall degrading enzymes. Accordingly, plant resistance to infection by broad host-range necrotrophs such as pectobacteria is quantitative, and it requires many genes to confer resistance. Perception of P/DAMPs by pattern recognition receptors (PRRs) on extracellular surfaces of plant cells in the apoplastic space leads to the induction of typical PTI responses such as ethylene/ jasmonate hormone biosynthesis and cell wall modifications, resulting in inhibition of disease proliferation (Mengiste, 2012). However, when invading pathogens successfully suppress PTI, by injection of effector proteins directly into plant cells, ETI is activated, wherein, recognition of effectors in a gene-for-gene defense pathway leads to a hypersensitive response (HR) and cell death at infection sites resulting in disease resistance (Jones and Dangl, 2006). Induction of PTI or ETI activates mitogen activated protein kinases (MAPKs) for signal transduction and regulation of downstream pathogen responsive genes involved in plant resistance to pathogen attack (Zhang and Klessig, 2001). PTI and ETI signals

converge in the MAPK cascade pathways and generally give rise to similar downstream responses. However, PTI is mostly effective against necrotrphic pathogens and PTI-related downstream responses include defense gene activation (Dodds and Rathjen, 2010;Meng and Zhang, 2013).

We previously reported on a potato cultivar *S. tuberosum* cv BP1 that shows significant tolerance to *P. c. brasilense* strain 1692 *(Pcb1692)* compared to the more susceptible *S. tuberosum* cv Valor (Kubheka et al., 2013). Thus, we wanted to use this tolerant vs susceptible model to further dissect the molecular basis of tolerance in *S. tuberosum* cv BP1, particularly in the early stages of infection (0 – 72 hpi) that signify the transition from asymptomatic to symptomatic phase in *Pcb1692* within the susceptible cultivar. Hence, in this study, we employed a time-course RNA-seq analyses to unravel the defense response in these potato cultivars during stem based colonization and infection by *Pcb1692.* The RNA-seq analysis allowed us to identify 6,139 and 8,214 DEGs in cultivars ‘BP1’ and ‘Valor’, respectively, compared to mock-inoculated controls, in the time-course. Expression profiles of the differentially expressed genes and gene ontology enrichment analysis revealed that the MPK3/MPK6 cascade, WRKY33 transcription factor and downstream defense genes including secondary wall biosynthetic genes are probable key components in the potato defense responses to *P. c. brasiliense*.

2.3 Results

2.3.1 Illumina sequencing and reads assembly

Stems of two potato cultivars, *S. tuberosum* cv BP1 (tolerant cultivar) and *S. tuberosum* cv Valor (susceptible cultivar) were inoculated with *P. carotovorum* subsp. *brasiliense* strain 1692 (*Pcb1692*) and samples collected at 0 (mock-inoculated), 6, 12, 24 and 72 hpi. In total, 30 RNA samples (comprising three biological replicates) from stem tissues of these two potato cultivars were obtained from the five time points and subjected to RNA-seq. Approximately 1.4 billion paired-end reads were generated in the time-course, producing an average of 46 million mapped reads per sample (Supplementary Table S2.1). In addition, over 80% of these reads could be mapped to the *S. tuberosum* reference sequence (*S. tuberosum* group Phureja DM1-3 516 R44), and approximately 92% were uniquely mapped (Supplementary Table S2.1). Furthermore, a principal component analysis (PCA) using transformed read counts from each biological replicate at each time-point, showed congruence of RNA-seq data among biological replicates, as well as, indicating that strong variation in our dataset was mainly due to the type of cultivar (i.e., inherent differences between the tolerant and susceptible cultivars) (Supplementary Fig. S2.1). In addition, some of the variation was explained by time [i.e., in the tolerant cultivar, main difference was observed at 12hpi, compared to mock-inoculated controls (0hpi), and at 72hpi in the susceptible cultivar compared to 0hpi] (Supplementary Fig. S2.1).

The current potato genome has 38,982 predicted gene models (PGSC_DM_v4.03; http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml) (Consortium, 2011). In this study, we identified expression of 38,688 genes by merging together transcripts reconstructed from each sample using Cufflinks software tool (v2.11) (Trapnell et al., 2012). Thus, the majority of annotated potato genes were detected (~ 99%). In addition, we identified 1,828 candidate novel protein coding expressed loci, present in both cultivars (Supplementary Table S2.2), based on the pipeline outlined in (Kwenda et al., 2016). These putative novel transcripts

represent potentially new information for improvement of the current potato genome annotation.

2.3.2 Pairwise comparisons of transcriptional dynamics between the susceptible and tolerant cultivar over time

Transcriptome profiling revealed a total number of 4718, 4503, 7577, 3505, and 5081 differentially expressed genes (DEGs) between *S. tuberosum* cv BP1 and *S. tuberosum* cv Valor at 0, 6, 12, 24 and 72 hpi, respectively (Supplementary Table S2.3). Transcriptional dynamics highlighting specific numbers of DEGs at individual time-points between these two cultivars, are shown in Fig. 2.1a. These comparisons, over time, revealed an exponential increase of DEGs in the tolerant cultivar in the early hours of infection [0-12 hpi (Fig. 2.1a)]. To investigate the functionality of the genes activated in response to *Pcb1692* infection in the tolerant cultivar compared to the susceptible cultivar, GO enrichment analysis was performed using g:Profiler web server (http://biit.cs.ut.ee/gprofiler/) against *S. tuberosum* ontologies (Reimand et al., 2016). Because of the large array of datasets over-represented under the three gene ontology categories namely; molecular function (MF), cellular component (CC) and biological process (BP), focus was only given to the BP responses in this study. The most overrepresented BP terms are shown in Fig. 2.2.

2.3.4 Transcriptional profiles in response to *Pcb1692* infection

To understand the transcriptional changes per cultivar in response to *Pcb1692* inoculation, cultivar-specific expression profiles were determined by comparing inoculated samples to their respective mock-inoculated samples (at 0 hpi). Towards this end, 6,139 and 8,214 DEGs were identified in the tolerant and susceptible cultivars, respectively (Fig. 2.1b,c). We found that the number of DEGs was initially higher in the susceptible cultivar (2,754 DEGs; 971 up-regulated and 1,783 down-regulated) at 6 hpi, compared to only 1,014 DEGs in the tolerant cultivar (684 up-regulated, 330 down-regulated) at this time point (Fig. 2.1b). This possibly reflects a faster rate of pathogen proliferation in the susceptible cultivar resulting in increased responses in the host. However, a marked increase in the number of DEGs in the tolerant

cultivar was observed at 12 hpi (up to 2-fold increase in the number of DEGs compared to 6 hpi; Fig. 2.1b). Furthermore, the highest number of DEGs was observed at 12 hpi in the tolerant cultivar (4,227 total DEGs; 2,265 up-regulated, 1,962 down-regulated) (Fig. 2.1b). Even though a peak in DEGs was observed at 12 hpi in the tolerant cultivar, the number of DEGs dropped significantly at 24 and 72 hpi. On the contrary, a marked increase of DEGs was observed at the later stages of infection (72 hpi) in the susceptible cultivar, and the highest number of DEGs was observed at this time-point (72hpi) (4,732 total DEGs; 2,455 up-regulated and 2,277 down-regulated). Together, these differences in expression profiles between the tolerant and susceptible cultivars, indicate differences in defense responses in these two cultivars.

2.3.5 The tolerant and susceptible cultivars employ similar sets of genes involved in pathogen recognition and wounding response

Among the identified DEGs, 4,210 were present in both cultivars in the time course (Fig. 2.1c,d and Supplementary Table S2.4). Among these were membrane localized receptor like kinases (RLKs) including FLAGELLIN SENSING 2 (PGSC0003DMG400008296, FLS2), EF-Tu receptor (EFR, PGSC0003DMG400023283), Wall-associated kinases (e.g., WAK1) and Brassinosteroid insensitive 1-associated kinases (BAK1) (Table 2.1). FLS2 and EFR are key pattern recognition receptors (PRRs) that recognize the conserved bacterial flagellin and EF-Tu proteins, respectively, thus triggering PTI defense signaling pathways in plants (Gómez-Gómez and Boller, 2000;Boller and Felix, 2009). The large number of RLKs present in both cultivars, suggests that these two cultivars employ fundamentally similar sets of pathogen recognition genes (Supplementary Table S2.4). Other DEGs related to plant defense responses were identified in both cultivars, including NAC domain-containing proteins and cytochrome P450 genes (e.g., PGSC0003DMG400030413) (Supplementary Table S2.4). Furthermore, genes involved in pathogen perception and response to wounding were differentially expressed in both cultivars. These included signaling genes encoding transcription factors such as MYB, WRKY, AP2 (*e.g.*, AP2-EREBP,

PGSC0003DMG400002272) and ethylene response factors (e.g. PGSC0003DMG400041451, ERF1) (Table 2.1 and Supplementary Table S2.4). These transcription factor families represent some of the major regulators of plant immune response pathways against necrotrophs (Lai and Mengiste, 2013). Additional wound responsive DEGs included RBOHD, lipoxygenases (e.g., LOX1, PGSC0003DMG400010859, NAC domain-containing proteins (*e.g.*, NAC002, PGSC0003DMG400032555), *JAR1* (PGSC0003DMG400033879), and *JAZ10* (PGSC0003DMG400006480) (Table 2.1).

Interestingly, in addition to PTI related responses, several DEGs encoding R-proteins that predominantly contain a nucleotide-binding site (NBS) and/or leucine-rich repeat (LRR) domain, were differentially expressed in both cultivars at 12, 24 and 72 hpi (Supplementary Table S2.4). Among these included resistance genes encoding R-proteins containing the coiled-coil (CC)-NBS-LRR and Toll interleukin 1 receptor (TIR)-NBS-LRR motifs (Table 2.1). Of these, PGSC0003DMG400001756 gene was up-regulated (~4-Fold) while PGSC0003DMG400008185 (CC-NBS-LRR) and PGSC0003DMG400013627 (TIR-NBS-LRR) were down-regulated (~2.9-Fold decrease) at 12 hpi compared to mock-inoculated samples of each cultivar (Table 2.1). Finally, different genes encoding CC-NBS- LRR, NBS-LRR, and CC-NBS-LRR motifs were differentially expressed at 72 hpi, with one *R*-protein encoding gene (PGSC0003DMG400003353) showing significant up-regulation (slightly over 32-Fold increase) in both cultivars (Table 2.1). *R*-gene mediated resistance leads to effector triggered immunity (ETI), a defense response which recognizes bacterial effector proteins. The actual role that the induced *R*-genes play in the two potato cultivars' response to *Pcb1692* is still unclear. Generally, ETI defense responses are not directly effective against necrotrophic pathogens such as *Pectobacterium* (Jones and Dangl, 2006).

2.3.6 Cultivar-specific transcriptional changes following inoculation with *Pcb1692*

Despite the high number of DEGs present in both cultivars (4,210 DEGs), only 1,929 and 4,004 DEGs were specific to the tolerant and susceptible cultivars, respectively (Fig. 2.1c, d and Supplementary Table S2.5). Among the 1,929 DEGs specific to the tolerant cultivar, GO

enrichment analyses using g:Profiler webserver revealed that 149 DEGs were overrepresented in the phosphorylation and negative regulation of cell death GO biological process categories (Fig. 2.3a and Supplementary Table S2.6). Interestingly, included in this category were defense-related signal transduction genes including MPK3 (PGSC0003DMG400030058), key in the activation of plant responses to biotic stress, and MPK4 (PGSC0003DMG401000057) which plays essential roles in pathogen defense signaling (Meng and Zhang, 2013). MPK3 was only induced in the tolerant cultivar in the early stages following infection (at 6, 12 and 24 hpi; ~5.7-Fold) but not in *S. tuberosum* cv Valor (Fig. 2.4a). Plant MAPK cascades are involved in the early transduction of perceived signals from PRRs activating a wide array of downstream defense responses, thus, playing a pivotal role in PTI. Additionally, MPK4 was up-regulated at 12 hpi in cultivar 'BP1' (2.3-Fold) but not in 'Valor'. Furthermore, defense-related transcription factors (TFs) such as WRKY-like transcription factor (PGSC0003DMG400011633, AtWRKY33) were enriched in this category. PGSC0003DMG400011633 (ortholog of AtWRKY33) was up-regulated in the tolerant cultivar by over 15 fold at 6, 12 and 24 hpi (Fig. 2.4a). WRKY33 plays key roles in the activation of downstream defense genes. Conversely, DEGs specific to the susceptible cultivar were associated with biological processes such as cell wall biogenesis, regulation of cellular component organization, and cellular response to DNA damage stimulus (Fig. 2.3b and Supplementary Table S2.6). Strikingly, genes overrepresented in the 'plant-type secondary cell wall biogenesis' comprising mainly secondary wall biosynthetic genes, were mainly down-regulated in the susceptible cultivar. Among these genes were cellulose synthases (e.g., *CESA4*, *CESA8* and *FRA8*; PGSC0003DMG400003822, PGSC0003DMG400028426, and PGSC0003DMG400000411 respectively), lignin biosynthesis genes (e.g. *IRX3* and *IRX9*; PGSC0003DMG400011148 and PGSC0003DMG400001769) and NAC domain-containing proteins (e.g., PGSC0003DMG400012113, respectively) involved in regulation of secondary wall biosynthesis (Fig. 4b and Supplementary Table S2.6). Furthermore, MYB83 (PGSC0003DMG400006868, MYB20), which regulates these secondary wall biosynthetic genes, was also associated with the 'plant-type secondary cell wall biogenesis' category and

down-regulated in the susceptible cultivar (Fig. 2.4b and Supplementary Table S2.6). Interestingly, these secondary wall biosynthetic genes were up-regulated in the tolerant cultivar when compared directly to the susceptible cultivar at each time-point (Fig. 2.4b). Thus, the up-regulation of these genes in the tolerant cultivar following *Pcb1692* infection could imply that they are possibly defense-related genes enhancing resistance to *Pcb1692*.

Furthermore, additional cultivar specific DEGs which potentially contribute to the compatibility or incompatibility of the host-pathogen interaction between potato and *Pcb1692* were also identified in the time-course. Among these were genes up-regulated only in the tolerant cultivar, including genes important in the ethylene biosynthesis and signaling pathway such as *ACS4* (PGSC0003DMG400021651, ~19.7-fold), *ACO* homolog (PGSC0003DMG400017190, ~4.3-fold), *EBF1* (PGSC0003DMG400015853, ~ 2-fold), *ERF1A* (PGSC0003DMG400010750, ~2.9-fold), and *EIL3* (PGSC0003DMG400021381, ~2.3-fold); as well as PGSC0003DMG400016769 (~5.5-fold) and PGSC0003DMG400008337 (MYB21, ~6.7-fold), homologs of WRKY33 and MYB63, important in regulation of defense responses and secondary cell wall biogenesis, respectively (Supplementary Table S2.5). In the susceptible cultivar, genes involved in the ethylene biosynthetic process such as ACC synthases and oxidases; *ACS9* (PGSC0003DMG400021426), *ACO4* (PGSC0003DMG400016714), respectively, were down-regulated (Supplementary Table S2.5). Additionally, susceptibility-related genes against necrotrophic pathogens such as MYC2 basic helix-loop-helix-leucine zipper (bHLH) transcriptional factors (e.g. PGSC0003DMG400007010, PGSC0003DMG400012237) were differentially expressed in the susceptible cultivar, between 12 and 72 hpi (Supplementary Table S2.5). Induction of MYC2 tends to enhance susceptibility to necrotrophic pathogens (Glazebrook, 2005).

2.3.7 Identification and functional characterization of novel genes

In addition to identifying expression profiles of known genes in the potato genome that are induced by *Pcb1692* infection, we also uncovered novel protein-coding potato transcripts responsive to *Pcb1692* inoculation. Strand-specific RNA-seq was used to identify a total of

1,828 novel CDS gene candidates assembled from reads mapped to intergenic regions using Cufflinks tool (as outlined in Chapter 3). In the present study, these candidate novel transcripts were assessed for their involvement in potato defense responses based on differential expression between cultivars 'Valor' and 'BP1'. Comparison of *Pcb1692*-inoculated samples to the mock-inoculated controls in each cultivar showed the highest number of DE novel candidate genes was at 12 hpi (549 and 511 in *S. tuberosum* cv Valor and BP1, respectively) (Supplementary Table S2.7). Only novel gene candidates showing statistically significant differential expression (adjusted p-value < 0.1) were considered. Furthermore, by using InterProScan5 (v5.11-51) (Jones et al., 2014), we characterized 28 (including 17 domains) and 32 (including 15 domains) candidate novel transcripts from cultivar 'BP1' and 'Valor', respectively (Supplementary Table S8). The GO terms assigned to each novel CDS transcript were visualized using WEGO (Web Gene Ontology Annotation Plot, http://wego.genomics.org.cn/cgi-bin/wego/index.pl). Candidate novel CDS genes were associated with defense related GO terms, including response to stress and immune response (Novel244, Novel1477, Novel2785 and Novel2787); and response to stimulus (Novel1477, Novel2785, Novel2477, Novel2724, Novel1946, Novel244 and Novel2787) (Fig. 2.5).

2.3.8 Validation of DEGs and novel candidates

To confirm the time-course RNA-seq data, five DEGs were randomly selected representing genes differentially expressed throughout the time-course in one or both cultivars. The expression profiles of these genes were validated experimentally using RT-qPCR. The RT-qPCR results were in agreement with the RNA-seq expression patterns (Fig. 2.6). In addition, eight of the1,828 novel CDS gene candidates were validated using RT-PCR (Supplementary Fig. S2.2).

2.4 Discussion

To our knowledge, this is the first transcriptome-wide study unraveling responses to *P. carotovorum* subsp. *brasiliense* infection in potato stems. Blackleg is an important disease of potato plants mainly in the field. It is caused mainly by members of the genus *Pectobacterium* such as *P. atrosepticum* and the emerging *P. c. brasiliense*. *Pectobacterium* species are usually found in tuber lenticels, on roots or colonizing and occluding potato plant xylem (Charkowski, 2015). Previous work on the host-pathogen interaction between potato and soft rot bacterial pathogens has mainly focused on the pathogen, that is, its pathogenicity and colonization patterns of tubers, roots and stems (Czajkowski et al., 2010;Kubheka et al., 2013). However, not much has been reported on the host potato stem responses against soft rot pathogens. Thus, this study provides new and relevant insights into stem-based defense mechanisms employed by potato plants during colonization and infection of xylem vessels by *P. c. brasiliense*. Therefore, in order to understand potato stem transcriptome dynamics elicited by *P. c. brasiliense* inoculation, we investigated differential gene expression following infection by this pathogen using time-course RNA-seq in tolerant and susceptible potato cultivars. A total of 4718, 4503, 7577, 3505, and 5081 DEGS were identified at 0, 6, 12, 24 and 72 hpi respectively (Fig. 2.1a and Supplementary Table S2.3), in pairwise comparisons between cultivars ‘Valor’ and ‘BP1’. The near exponential increase in up-regulated DEGs in ‘BP1’ induced by *Pcb1692* immediately following inoculation is suggestive of an early activation of defense responses in this tolerant cultivar. Furthermore, the highest number of DEGs was observed at 12 hpi in ‘BP1’ when comparing inoculated samples from each cultivar to mock-inoculated control samples or in pairwise comparisons with ‘Valor’. This implies that 12 hpi could be a key time-point contributing to the subsequent tolerance in ‘BP1’ (Fig. 2.1a,b). Collectively, these results imply that type of cultivar (in this case ‘BP1’) has a significant role in defining the early defense responses to bacterial pathogen attack, and these early defenses lead to overall tolerance or susceptibility, resulting in a compatible or incompatible interaction with *Pcb1692* at the later stages of infection (72 hpi and beyond).

2.4.1 Pathogen-recognition and signal transduction genes regulated by *Pcb1692* infection

Plants possess pattern recognition receptors (PRRs) which perceive conserved molecular signatures of invading pathogens called PAMPs or recognize signals arising from damage inflicted on the plant by pathogens (DAMPs) in the extracellular environment. Recognition of D/PAMPs initiates plants basal immunity, termed PTI (Dodds and Rathjen, 2010). Generally, PTI defense responses do not involve hypersensitive response (HR) cell death, making PTI important against necrotrophic pathogens. PRRs belong to classes of receptor-like kinases (RLKs). In the present study, several leucine-rich repeats (LRR) RLKs shared between the tolerant and susceptible cultivar were identified (Table 2.1 and Supplementary Table S2.2). Included among these are the well-characterized plant PRRs, such as FLS2, which recognizes the bacterial flagellin conserved peptide (flg22) and homologs of the Arabidopsis EFR receptor which recognizes bacterial EF-Tu (elf18) (Zipfel et al., 2006). Induction EFR genes was much higher in the tolerant cultivar compared to the susceptible cultivar (Table 2.1). Expression of FLS2, was upregulated in both cultivars up to 12 hpi (Table 2.1). Perception of bacterial flagellin and EF-TU triggers early defense responses in plants including strong activation of MPK3/MPK6 and MPK4 cascades, ethylene biosynthesis and reactive oxygen species (ROS), which in turn signal downstream defenses such as cell wall strengthening (Meng and Zhang, 2013).

Throughout the time-course WAK receptor genes which perceive DAMPs due to the action of cell wall degrading enzymes, were modulated following inoculation with *Pcb1692* in both cultivars. Some WAKs were up or down-regulated in both cultivars (Supplementary Table S2.4). However, most of the WAK1 genes were up-regulated in the tolerant cultivar (e.g. PGSC0003DMG400011792), when comparing inoculated samples from each cultivar to mock-inoculated controls (Table 2.1). Induction of WAKs has been associated with perception of oligogalacturonides (OGs) and bacterial EF-Tu in defense responses against necrotrophic pathogens (Mengiste, 2012). Thus, the observed up-regulation of WAK1 in BP1 correlates

with enhanced pathogen perception. Another crucial RLK is BAK1 which interacts with and forms complexes with PRRs including FLS2, immediately upon perception of D/PAMPs thereby linking the perceived cues with innate immune responses through activation of the MAPK signaling cascades (Mengiste, 2012;Meng and Zhang, 2013). Four genes encoding BAK1 RLKs were differentially expressed in both or one of the cultivars in response to *Pcb1692* infection (Supplementary Table S4). Three BAK1 encoding genes were up-regulated only in the tolerant cultivar, compared to the susceptible cultivar (Supplementary Table S5). BAK1 is central in PTI immunity regulation and Arabidopsis *bak1* mutants have higher susceptibility to necrotrophic pathogens (Dodds and Rathjen, 2010). Collectively, these results emphasize the congruence of pathogen recognition in the two cultivars, although, the higher pathogen-induced expression of WAK1 and BAK1 in 'BP1' possibly contributes to the strong defense response and observed tolerance in cultivar 'BP1'.

Transduction of signals perceived by PRRs and BAK1 complexes is mediated by plant MAPK pathways which transfer signals to downstream components of host immunity. Typically, MAPK cascades comprise MAPK kinase kinase (MAPKKK) which receive signal from PRR/BAK1 complexes. Their activation in turn regulates MAPK kinase (MAPKK) which phosphorylates downstream MAPKs. The MAPK cascades are involved in PTI and ETI and they regulate downstream activities of various substrates including transcription factors (Dodds and Rathjen, 2010). In Arabidopsis flagellin perception can activate two independent MAPK cascade pathways, the MAPKKK-MAPKK (MPKK4/MPKK5)-MPK3/MPK6 cascade and the MAPKKK-MPKK1/MPKK2-MPK4 cascade leading to downstream activation of early defense response genes including WRKY22/29 and WRKY33 transcription (Asai et al., 2002;Meng and Zhang, 2013). In this study, five MAPKKKs (PGSC0003DMG400028666, PGSC0003DMG400018992, PGSC0003DMG400024820, PGSC0003DMG400015448 and PGSC0003DMG400022210) were significantly up-regulated at one or more time-points in one or both cultivars, one MAPKK (, PGSC0003DMG400033696) was up-regulated in 'Valor' at 72 hpi. Remarkably, two MAPKs, MPK3 (PGSC0003DMG400030058, AtMPK3) and MPK4

(PGSC0003DMG401000057), critical in flg22-PTI immune responses were only up-regulated in BP1 at 6, 12, and 24 hpi. MPK3 and MPK4 play essential roles in signaling pathogen-induced plant disease resistance, by activation of WRKY33 and WRKY22 in PTI-related defense responses (Asai et al., 2002). In Arabidopsis, MPK4 represses salicylic acid (SA)-dependent resistance (which often result in HR-cell death and are important in defenses against biotrophs) and it interacts with intermediate substrates which activate WRKY33 downstream (Andreasson et al., 2005). Strikingly, WRKY33 was up-regulated only in the tolerant cultivar throughout the time-course in response to *Pcb1692* infection. Taken together, these findings suggest that the early and persistent induction of MPK3/MPK4 cascade genes and the activation of downstream transcription factors in the tolerant cultivar enhance transduction of the perceived stress stimuli leading to stronger cellular defense responses in 'BP1' against *P. c. brasiliense* challenge.

2.4.2 Transcription factors responsive to *P. carotovorum* subsp *brasiliense* infection

We identified DEGs representing four families of transcription factors (MYB, MYC2 (bHLH), AP2/ERF, and WRKY) modulated in response to *Pcb1692* infection. Timely regulation and coordinated expression of genes in plant immune response signaling pathways is central to effective defense against pathogens (Mengiste, 2012). Transcription factors connect pattern recognition receptors (PPR) perception and MAPK signaling to downstream gene expression. Many families of transcription factors such as ERFs, MYBs and WRKYs are involved in immunity to necrotrophic pathogens (Lai and Mengiste, 2013). For instance, in this study, WRKY33 associated with plant disease resistance was only up-regulated in the tolerant cultivar in response to *Pcb1692* challenge. WRKY33, a pathogen-inducible transcription factor, was constitutively expressed throughout the time-course (~22.1-fold induction) in the tolerant cultivar. WRKY33 impacts significantly immune responses to necrotrophs. Overall, WRKY33 activates cellular responses downstream of MPK3/MPK6 and MPK4 in PTI immune signaling induced by bacterial flagellin (Asai et al., 2002;Meng and Zhang, 2013).

Defense-related ERFs integrate signals from jasmonate and ethylene pathways in order to transcriptionally activate plant defense responses to necrotrophs (Mengiste, 2012). Here, ERF1 genes were significantly expressed in both cultivars at all the time-points except at 6 hpi in the susceptible cultivar. DEGs for ERF1 were more induced in BP1 compared to Valor in the time-course. ERF1 positively regulates plant resistance to necrotrophs. However, some ERF genes such as ERF4 and ERF5, associated with susceptibility to necroptrophs were also identified and were differentially expressed in the susceptible cultivar following inoculation with *Pcb1692*. The homolog of *AtERF2*, PGSC0003DMG400026261, was induced in both cultivars at 6 hpi but specifically expressed in the tolerant cultivar at 12 and 24 hpi and in the susceptible cultivar at 72 hpi. Interestingly, ERF2 expression patterns showed a 6-fold increase in BP1 at 12 hpi when compared to Valor, indicative of an enhanced and stronger defense response in the tolerant cultivar. In Arabidopsis, *AtERF4* negatively regulates expression of jasmonate responsive defense genes and resistance to necrotrophs. In contrast, *AtERF2* is a positive regulator of defense genes in the jasmonate signaling pathway, conferring resistance to necrotrophic pathogens (Grennan, 2008).

Other differentially expressed transcription factors belong to MYB and MYC families. MYC2 is associated with repression of responses to necrotrophic pathogen infection (Mengiste, 2012) and was mostly down-regulated throughout the time-course in both cultivars, except at the later stages of infection (72 hpi) in the susceptible cultivar when blackleg symptoms are evident. The MYB transcription factor, MYB83 (PGSC0003DMG400006868) was repressed in the susceptible within the time-course, but was up-regulated in the tolerant cultivar (when using pairwise comparisons between cultivars ‘Valor’ and ‘BP1’ at each time-point). MYB83 is a close homolog of and acts redundantly with MYB46. MYB46/MYB83 transcription factors are master regulators of secondary cell wall formation in Arabidopsis, directly regulating expression of genes involved in lignin, cellulose and hemicellulose biosynthesis, including

among others, *SND1*, *CESA4*, CESA7, *CESA8*, MYB58, MYB56 and MYB63 (McCarthy et al., 2009;Ko et al., 2014). Thus, MYB83 appears to be a key component of cell wall modifications in downstream (late) plant defense responses.

2.5 Conclusion

In this study, we presented the first time-course RNA-seq analysis focusing on potato stem-based defense responses to *P. c. brasiliense* attack. Our findings suggest that differential regulation and expression of PTI-related genes play a central role in cultivar 'BP1' pathogen induced defense responses. In addition, by detecting cultivar specific DEGs, we identified gene sets that distinguished the tolerant and susceptible cultivars. Thus the type of cultivar has a role in plant resistance to *Pectobacterium* infection. Furthermore our time-course data showed induction of defense-related genes at different time-points and stronger expression of majority of these genes in the tolerant cultivar. The highest number of DEGs was identified at 12 hpi in the tolerant cultivar, suggesting that key defense mechanisms are regulated early against *P. c. brasiliense* challenge.

2.6 Materials and Methods

2.6.1 Plant material and RNA preparation

Seed tubers of *Solanum tuberosum* cv. Valor *S. tuberosum* cv. BP1, susceptible and tolerant to *P. carotovorum* subsp *brasiliense* strain 1692 (*Pcb1692*) infection, respectively, were greenhouse grown under standard conditions (22 to 26 °C, 16 h light/ 8 h dark photoperiod and 70% relative humidity). Stem inoculations were done following the approach previously described in (Kubheka et al., 2013), except that in this study we only used wild-type *P. carotovorum* subsp. *brasiliense* 1692 for the inoculations. Inoculated plants were assessed and sampled, within 2 cm above or below the point of inoculation, at 0, 6, 12, 24, and 72 hours post inoculation (hpi) in triplicates (three plants were pooled together for each biological replicate). Samples at 0 hpi were mock-inoculated with $MgSO_4$ buffer and used as controls. Total RNA was extracted from individual time-points and replicates independently using the QIAGEN RNeasy plant mini kit (Qiagen) including DNAse treatment (Qiagen). RNA was quantified using the NanoDrop (Thermo Scientific, Sugarland, TX, USA) and the quality and integrity checked using Agilent 2100 BioAnalyzer system (Agilent, Santa Clara, CA, USA).

2.6.2 cDNA library construction and Illumina sequencing

The construction cDNA libraries and sequencing were carried out at the Beijing Genomics Institute (BGI-Shenzhen, China; http://www.genomics.cn/en/index). The quality of total RNA samples from individual biological replicates (n=3) from each time-point was assessed using Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit) and NanoDrop, and 200 ng aliquots were used for poly(A) mRNA isolation and preparation of cDNA libraries using the TruSeq RNA sample Prep Kit v2 (Illumina, San Diego, CA, USA) following manufacturer's instructions . The libraries were quality checked and quantified using Agilent BioAnalyzer 2100 system and qPCR. Finally, the libraries were sequenced with an Illumina HiSeq 2000 sequencer generating 90 bp paired-end reads.

2.6.3 Data access

The data have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through the GEO accession number, GSE74871.

2.6.4 Differential expression analysis

Clean paired-end reads from each time-point in each cultivar were initially quality checked using FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) and mapped to the potato reference genome using TopHat2 (version 2.0.13) (Kim et al., 2013). Transcript reconstruction was done using Cufflinks software tool (version 2.2.1) (Trapnell et al., 2012). HTSeq-count (Anders et al., 2014) and DESeq2 package (Love et al., 2014) were used to make read counts and perform a time-series differential expression analysis, respectively. A False Discovery Rate (FDR) threshold of 10% and an absolute log2 fold change > 1 were used to determine differentially expressed genes.

2.6.5 Gene ontology enrichment analysis

Functional enrichment analysis of differentially expressed genes obtained from each comparison (direct pairwise comparison between cultivars 'Valor' and 'BP1' or cultivar specific comparisons of inoculated samples from each time-point to mock-inoculated controls) was performed using g:Profiler web server (Reimand et al., 2016).

2.6.6 Orthology detection

Orthology detection was performed using BLASTp searches to compare sequences of differentially expressed genes to the Arabidopsis TAIR genome using the ProteinOrtho software (Lechner et al., 2011), with the default cutoff E-value: 1.0E-05. Subsequently TAIR annotations were used in Figures and Tables in the text.

2.6.7 RT-qPCR validation of RNA-seq data

For RT-qPCR, first-strand cDNA synthesis was done from total RNA using Superscript III First-Strand cDNA Synthesis SuperMix kit (Invitrogen, USA) following manufacturer's instructions. Quantitative real-time PCR using Applied Biosystems SYBR Green Master Mix was performed in the QuantStudio 12K Flex Real-Time PCR system (Life Technologies, Carlsbad, CA, USA). For RT-qPCR, 2 µl of sample was added to 8 µl of Applied Biosystems SYBR Green Master Mix and primers at a concentration of 0.4 µM. The cycling conditions were as follows: an initial denaturation at 50 °C for 5 min and 95 °C for 2 min followed by 45 cycles of 95 °C for 15 s and 60 °C for 1 min. Each sample was run in triplicate. The samples were normalized to 18S rRNA and elongation factor 1-α (PGSC0003DMG400020772, ef1α) as the reference genes and the mock treated samples used as calibrators (Nicot et al., 2005). The comparative CT ($\Delta\Delta^{ct}$) method was used to measure relative expression (Livak and Schmittgen, 2001). Two tailed Student's *t*-test (unequal variances) was used to check whether RT-qPCR results were statistically different when comparing inoculated samples to mock-inoculated samples (**P<0.01; *P<0.05). Primers used were designed online using Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and are listed in Supplementary Table S2.9.

2.6.8 RT-PCR verification of candidate novel CDS transcripts

First-strand cDNA was synthesized as outlined in the RT-qPCR validation section above. The PCR was done on Bio-RAD T100TM Thermal Cycler end-point PCR (Bio-RAD, USA). The PCR reaction mix consisted of 12.5 µl KAPA HiFi HotStart Ready mix (2X), 0.5 µM of each forward and reverse primer, 1 µl template cDNA in a final reaction volume of 25 µl. PCR conditions were: 98°C for 3 min; 28 cycles of 98°C for 30 s, annealing for 60 s, 72°C for 90 s, and final extension at 72°C for 5 min. The PCR products were analyzed on 1.5 % agarose gel including 1 kb DNA molecular weight ladder (NEB, UK). All the primers were synthesized by Inqaba Biotech, South Africa (Supplementary Table S2.10).

2.7 Supplementary Data

Fig. S2.1: Principal component analysis (PCA) plot showing congruence of RNA-seq data among biological replicates and indicating that the main differences in the dataset are due to the type of cultivar (i.e., differences between the susceptible and tolerant cultivars) and also influenced by time (i.e., different time-points used in this study). The PCA plot was generated using read count data transformed using the regularized-logarithm transformation (rlog) function implemented in DESeq2 package (Love et al., 2014).

Fig. S2.2: RT-PCR validation of eight novel CDS candidates using agarose gel. Lane 1. 1 kb DNA ladder, Lane 2 and 3. Novel1253 in the tolerant and susceptible cultivars, respectively, Lane 4 and 5. Novel917 in the tolerant and susceptible cultivars, respectively, Lane 6, Novel806 in susceptible cultivar, Lane 7 and 9. Novel2142 in the susceptible cultivar, Lane 8. Novel1481 in the tolerant cultivar, Lane 10. Novel750 in the susceptible cultivar, Lane 11. Novel2049 in the susceptible cultivar, Lane 12. Novel2769 in the susceptible cultivar.

Table S2.1: Mapping statistics

Table S2.2: List of identified novel CDS gene candidates

Table S2.3: List of differentially expressed genes at each time-point between cultivars ‘Valor’ and ‘BP1’

Table S2.4: Common DE genes present in both ‘Valor’ and ‘BP1’ at each time-point compared to mock-inoculated samples

Table S2.5: Cultivar-specific differentially expressed genes

Table S2.6: Cultivar-specific DEGs overrepresented in various Gene Ontology processes

Table S2.7: Differentially expressed candidate novel CDS transcripts in the time-course

Table S2.8: Domains assigned to novel CDS candidates using InterProScan5 software tool.

Table S2.9: RT-qPCR primers used to validate RNA-seq data

Table S2.10: RT-PCR primers used for novel CDS candidate transcripts validation

2.8 References

ANDERS, S., PYL, P.T., & HUBER, W. 2014. HTSeq–A Python framework to work with high-throughput sequencing data. *Bioinformatics*, btu638.

ANDREASSON, E., JENKINS, T., BRODERSEN, P., THORGRIMSEN, S., PETERSEN, N.H.T., ZHU, S., QIU, J.-L., MICHEELSEN, P., ROCHER, A., PETERSEN, M., NEWMAN, M.-A., BJØRN NIELSEN, H., HIRT, H., SOMSSICH, I., MATTSSON, O., & MUNDY, J. 2005. The MAP kinase substrate MKS1 is a regulator of plant defense responses. *The EMBO Journal* 24, 2579-2589.

ASAI, T., TENA, G., PLOTNIKOVA, J., WILLMANN, M.R., CHIU, W.-L., GOMEZ-GOMEZ, L., BOLLER, T., AUSUBEL, F.M., & SHEEN, J. 2002. MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature* 415, 977-983.

BOLLER, T., & FELIX, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annual Review of Plant Biology* 60, 379-406.

CHARKOWSKI, A.O. 2015. Biology and control of Pectobacterium in potato. *American Journal of Potato Research* 92, 223-229.

CONSORTIUM, P.G.S. 2011. Genome sequence and analysis of the tuber crop potato. *Nature* 475, 189-195.

CZAJKOWSKI, R., DE BOER, W.J., VELVIS, H., & VAN DER WOLF, J.M. 2010. Systemic colonization of potato plants by a soilborne, green fluorescent protein-tagged strain of Dickeya sp. biovar 3. *Phytopathology* 100, 134-142.

DAVIDSSON, P.R., KARIOLA, T., NIEMI, O., & PALVA, E.T. 2013. Pathogenicity of and plant immunity to soft rot pectobacteria. *Frontiers in Plant Science* 4.

DE BOER, S., LI, X., & WARD, L. 2012. Pectobacterium spp. associated with bacterial stem rot syndrome of potato in Canada. *Phytopathology* 102, 937-947.

DODDS, P.N., & RATHJEN, J.P. 2010. Plant immunity: towards an integrated view of plant–pathogen interactions. *Nature Reviews Genetics* 11, 539-548.

DUARTE, V., DE BOER, S., WARD, L., & OLIVEIRA, A. 2004. Characterization of atypical Erwinia carotovora strains causing blackleg of potato in Brazil. *Journal of Applied Microbiology* 96, 535-545.

GAO, L., TU, Z.J., MILLETT, B.P., & BRADEEN, J.M. 2013. Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-Phytophthora infestans interactions. *BMC Genomics* 14, 340.

GLASNER, J., MARQUEZ-VILLAVICENCIO, M., KIM, H.-S., JAHN, C., MA, B., BIEHL, B., RISSMAN, A., MOLE, B., YI, X., & YANG, C.-H. 2008. Niche-specificity and the variable fraction of the Pectobacterium pan-genome. *Molecular Plant-Microbe Interactions* 21, 1549-1560.

GLAZEBROOK, J. 2005. Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual Review of Phytopathology.* 43, 205-227.

GÓMEZ-GÓMEZ, L., & BOLLER, T. 2000. FLS2: an LRR receptor–like kinase involved in the perception of the bacterial elicitor flagellin in Arabidopsis. *Molecular Cell* 5, 1003-1011.

GONG, L., ZHANG, H., GAN, X., ZHANG, L., CHEN, Y., NIE, F., SHI, L., LI, M., GUO, Z., & ZHANG, G. 2014. Transcriptome Profiling of the Potato (Solanum tuberosum L.) Plant under Drought Stress and Water-Stimulus Conditions. *PloS One* 10, e0128041-e0128041.

GRENNAN, A.K. 2008. Ethylene response factors in jasmonate signaling and defense response. *Plant Physiology* 146, 1457-1458.

GYETVAI, G., SONDERKAER, M., GOBEL, U., BASEKOW, R., BALLVORA, A., IMHOFF, M., KERSTEN, B., NIELSEN, K.-L., & GEBHARDT, C. 2012. The transcriptome of compatible and incompatible interactions of potato (*Solanum tuberosum*) with Phytophthora infestans revealed by DeepSAGE analysis. *PLoS One* 7, e31526.

JONES, J.D., & DANGL, J.L. 2006. The plant immune system. *Nature* 444, 323-329.

JONES, P., BINNS, D., CHANG, H.-Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., & NUKA, G. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30**,** 1236-1240.

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**,** R36.

KO, J.-H., JEON, H.-W., KIM, W.-C., KIM, J.-Y., & HAN, K.-H. (2014). The MYB46/MYB83-mediated transcriptional regulatory programme is a gatekeeper of secondary wall biosynthesis. *Annals of Botany***,** mcu126.

KUBHEKA, G.C., COUTINHO, T.A., MOLELEKI, N., & MOLELEKI, L.N. 2013. Colonization patterns of an mCherry-Tagged Pectobacterium carotovorum subsp. brasiliense Strain in potato plants. *Phytopathology* 103**,** 1268-1279.

LAI, Z., & MENGISTE, T. (2013). Genetic and cellular mechanisms regulating plant responses to necrotrophic pathogens. *Current Opinion in Plant Biology* 16**,** 505-512.

LECHNER, M., FINDEIß, S., STEINER, L., MARZ, M., STADLER, P.F., & PROHASKA, S.J. (2011). Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics* 12**,** 124.

LIVAK, K.J., & SCHMITTGEN, T.D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods* 25**,** 402-408.

LOVE, M.I., HUBER, W., & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15**,** 550.

MASSA, A.N., CHILDS, K.L., & BUELL, C.R. 2013. Abiotic and biotic stress responses in group Phureja DM1-3 516 R44 as measured through whole transcriptome sequencing. *The Plant Genome* 6.

MCCARTHY, R.L., ZHONG, R., & YE, Z.-H. 2009. MYB83 is a direct target of SND1 and acts redundantly with MYB46 in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant and Cell Physiology* 50**,** 1950-1964.

MENG, X., & ZHANG, S. 2013. MAPK cascades in plant disease resistance signaling. *Annual Review of Phytopathology* 51**,** 245-266.

MENGISTE, T. 2012. Plant immunity to necrotrophs. *Annual Review of Phytopathology* 50**,** 267-294.

NICOT, N., HAUSMAN, J.-F., HOFFMANN, L., & EVERS, D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany* 56**,** 2907-2914.

PANDA, P., FIERS, M., ARMSTRONG, K., & PITMAN, A. 2012. First report of blackleg and soft rot of potato caused by Pectobacterium carotovorum subsp. brasiliensis in New Zealand. *New Disease Report* 26**,** 15.

REIMAND, J., ARAK, T., ADLER, P., KOLBERG, L., REISBERG, S., PETERSON, H., & VILO, J. 2016. g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*.

TOTH, I.K., & BIRCH, P.R. 2005. Rotting softly and stealthily. *Current Opinion in Plant Biology* 8**,** 424-429.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D.R., PIMENTEL, H., SALZBERG, S.L., RINN, J.L., & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* 7**,** 562-578.

VAN DER MERWE, J.J., COUTINHO, T.A., KORSTEN, L., & VAN DER WAALS, J.E. 2010. Pectobacterium carotovorum subsp. brasiliensis causing blackleg on potatoes in South Africa. *European Journal of Plant Pathology* 126**,** 175-185.

WANG, Z., GERSTEIN, M., & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10**,** 57-63.

ZHANG, S., & KLESSIG, D.F. 2001. MAPK cascades in plant defense signaling. *Trends in Plant Science* 6**,** 520-527.

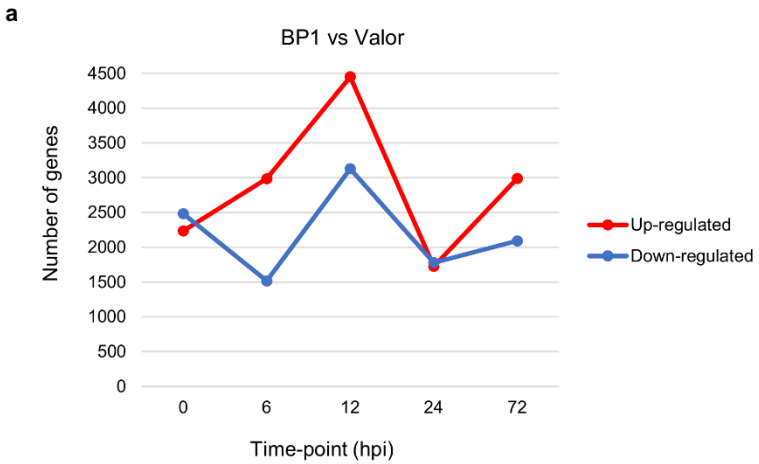
ZIPFEL, C., KUNZE, G., CHINCHILLA, D., CANIARD, A., JONES, J.D., BOLLER, T., & FELIX, G. 2006. Perception of the bacterial PAMP EF-Tu by the receptor EFR restricts Agrobacterium-mediated transformation. *Cell* 125**,** 749-760.

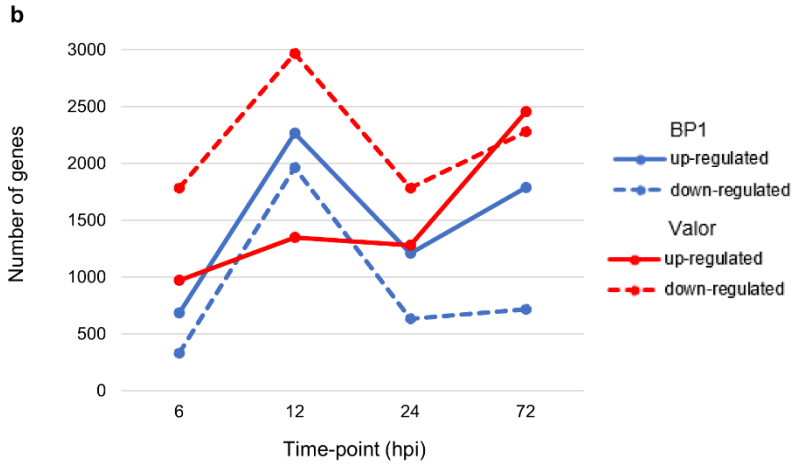
Table 2.1: Summary of selected pattern recognition receptors and intracellular receptors activated in response to *Pcb1692* infection

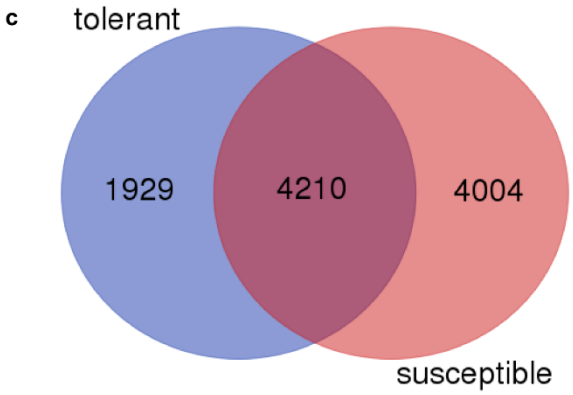
	6h[a]		12h		24h		72h			
Potato gene ID	**Valor**	**BP1**	**Valor**	**BP1**	**Valor**	**BP1**	**Valor**	**BP1**	**Arabidopsis ID**	**Gene name**
Pathogen recognition										
RLKs										
PGSC0003DMG400003195	6,47	4,54	-	12,46	-	3,34	6,49	2,35	AT3G05660	AtRLP33 (FLS2)
PGSC0003DMG400020697	-	-	6,59	2,58	-	-	3,25	7,06	AT5G46330	FLS2
PGSC0003DMG400020848	13,45	-	8,11	3,27	5,50	-	-	1,00	AT5G46330	FLS2
PGSC0003DMG400006502	-	3,32	3,18	5,82	2,66	6,06	11,88	5,58	AT3G47570	(EFR)
PGSC0003DMG400011932	-	-	4,03	-	-	-	3,32	2,69	AT3G47570	(EFR)
PGSC0003DMG400023283	5,58	6,11	-	44,32	7,31	9,71	6,77	12,21	AT5G20480	EFR
PGSC0003DMG400011792	0,21	5,54	-	15,78	-	11,79	-	10,20	AT1G21250	WAK1
PGSC0003DMG400025668	34,30	-	16,11	8,46	9,19	-	9,38	-	AT1G21240	WAK3
PGSC0003DMG400038918	8,17	-	-	4,47	-	-	-	-	AT2G23770	LYK4 (CERK1)
PGSC0003DMG401015527	19,48	6,38	-	11,92	6,80	4,61	7,79	-	AT2G19210	FRK1
PGSC0003DMG400003211	3,71	2,69	2,22	4,68	-	-	-	-	AT1G73080	PEPR1
PGSC0003DMG400020327	39,81	18,23	12,60	25,39	9,99	5,82	13,32	-	AT3G53810	LECRK42 (BAK1)
NBS-LRR										
PGSC0003DMG400001756	4,69	3,12	4,92	-	-	-	4,63	-	AT5G38344	
PGSC0003DMG400003353	8,57	33,82	-	-	-	62,25	37,27	76,11		
PGSC0003DMG400021469	4,03	-	-	-	2,93	2,71	3,39	2,33		
PGSC0003DMG400008185	0,45	-	0.52	0.34	-	0.44	-	-		
PGSC0003DMG400013627	0.04	-	0.41	0.41	-	-	-	-		
Transcription factors										
WRKY family										
PGSC0003DMG402006935	-	-	-	614,38	-	480,44	38,10	1225,39	AT5G15130	WRKY72
PGSC0003DMG400021895	8,24	17,93	5,50	63,46	-	51,15	24,28	12,08	AT5G13080	WRKY75

PGSC0003DMG400009103	40,22	51,07	-	38,91	3,89	30,02	-	27,51	AT5G24110	WRKY30
PGSC0003DMG400016441	10,66	6,66	-	15,07	-	-	17,22	6,83	AT1G62300	WRKY6
PGSC0003DMG400018081	20,02	13,72	-	44,88	7,30	6,68	46,75	15,32	AT5G15130	WRKY72
PGSC0003DMG400019824	4,64	9,24	-	14,10	-	9,14	6,38	-	AT1G80840	WRKY40
PGSC0003DMG400020206	22,38	1283646,84	-	17915245,67	-	7114329,79	398,92	2487951,03	AT3G01970	WRKY45
MYB family										
PGSC0003DMG400001504	-	9,74	-	-	5,99	-	-	7,24	AT3G24310	MYB305, ATMYB71
PGSC0003DMG400003890	-	-	-	7,60	2,50	6,10	6,08	7,56	AT5G60890	ATMYB34, ATR1, MYB34
PGSC0003DMG400004612	1348,14	-	-	623,99	-	-	-	-	AT1G48000	MYB112
PGSC0003DMG400005641	5,11	-	-	4,00	-	-	-	-	AT2G47190	ATMYB2, MYB2
PGSC0003DMG400011048	32,93	3,32	-	13,06	25,02	19,40	59,43	43,77	AT3G09600	Homeodomain-like superfamily protein
PGSC0003DMG401010883	18,23	-	-	11,39	10,53	30,57	14,17	38,47	AT3G46130	ATMYB48
PGSC0003DMG402004611	-	-	-	9,35	-	8,15	4,17	-	AT2G47190	ATMYB2
AP2/ERF family										
PGSC0003DMG400002272	163,24	130,09	-	284,39	-	251,38	54,69	-	AT5G47220	ATERF2
PGSC0003DMG400016812	6,02	-	-	2,29	-	-	3,57	-	AT3G16770	RAP2,3, ATEBP, ERF72, EBP
PGSC0003DMG400026260	90,51	33,98	-	42,87	-	15,53	42,94	6,12	AT4G17500	ATERF-1, ERF-1
PGSC0003DMG400014594	3,85	-	2,67	4,41	-	4,22	7,02	-	AT3G23240	ERF1, ATERF1
PGSC0003DMG400026046	162813,36	1025615,53	13345,41	22905524,55	-	30550103,67	1762294,29	302026,73	AT2G44840	ATERF13, EREBP, ERF13
PGSC0003DMG400026261	105,16	73,89	-	83,78	-	17,26	52,96	-	AT5G47220	ATERF2, ATERF-2, ERF2
Other wound-responsive genes										
PGSC0003DMG400024754	24,62	9,59	294,80	19,98	7,55	1,00	6,48	-	AT1G09090	ATRBOHB, ATRBOHB-BETA, RBOHB
PGSC0003DMG400010859	18,86	29,73	-	226,35	-	85,36	8,07	17,27	AT1G55020,1	LOX1
PGSC0003DMG400006480	-	-	-	8,31	-	-	171,49	-	AT5G13220,1	JAZ10, TIFY9, JAS1
PGSC0003DMG400001223	8,37	9,47	106,28	15,45	7,62	10,81	3,46	5,46	AT2G43000	anac042, NAC042
PGSC0003DMG400039898	8,84	21,38	40,66	24,12	8,56	26,52	7,76	20,92	AT2G43000	anac042, NAC042

[a]Fold changes in comparison to mock –inoculated controls







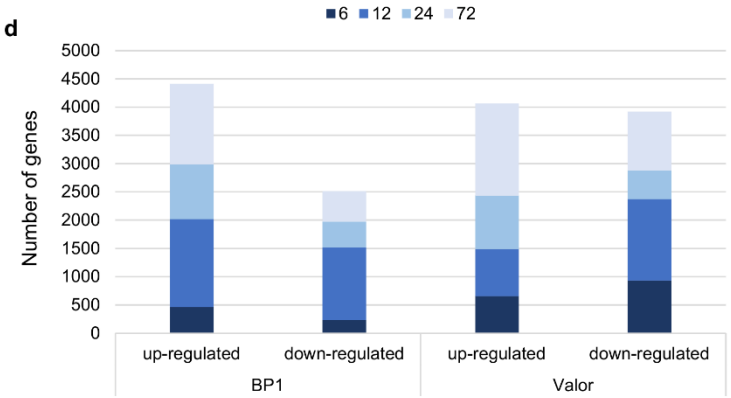


Fig. 2.1: (**a**) Pairwise comparison of DEGs in cultivars 'Valor' and 'BP1' in the time-course showing number of DEGs up- and down-regulated in cultivar 'BP1' compared to cultivar 'Valor'. (**b**) Cultivar specific DEGs between inoculated samples and mock-inoculated controls in each cultivar independently. (**c**) In total, 1,929 and 4,004 DEGs were identified and are specific to the tolerant and susceptible cultivar, respectively. Of these, 554 and 1137 DEGs in 'BP1' and 'Valor', respectively, represent intrinsic cultivar differences, and are related to plant growth and/or development (Table S5). In addition, 4,210 DEGs were present in both cultivars in the time-course. *Control group represents DEGs obtained between 'BP1' and 'Valor' at 0 h time-point. (**d**) Graph showing DEGs up- or down-regulated in both cultivars at individual sampling time-points (6, 12, 24 and 72 hpi).

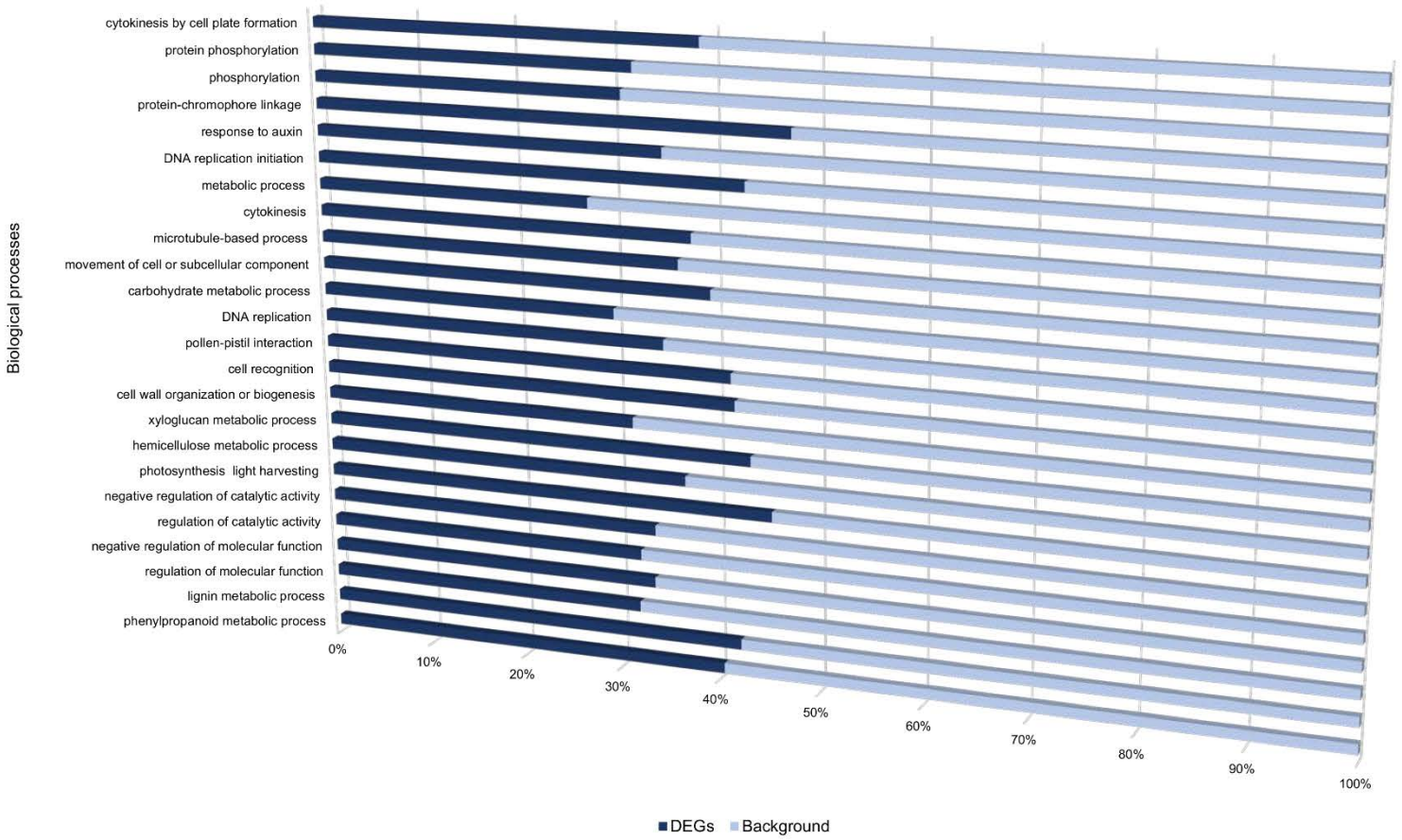


Fig. 2.2: Gene ontology biological processes overrepresented between cultivars ‘Valor’ and ‘BP1’ in the time-course.

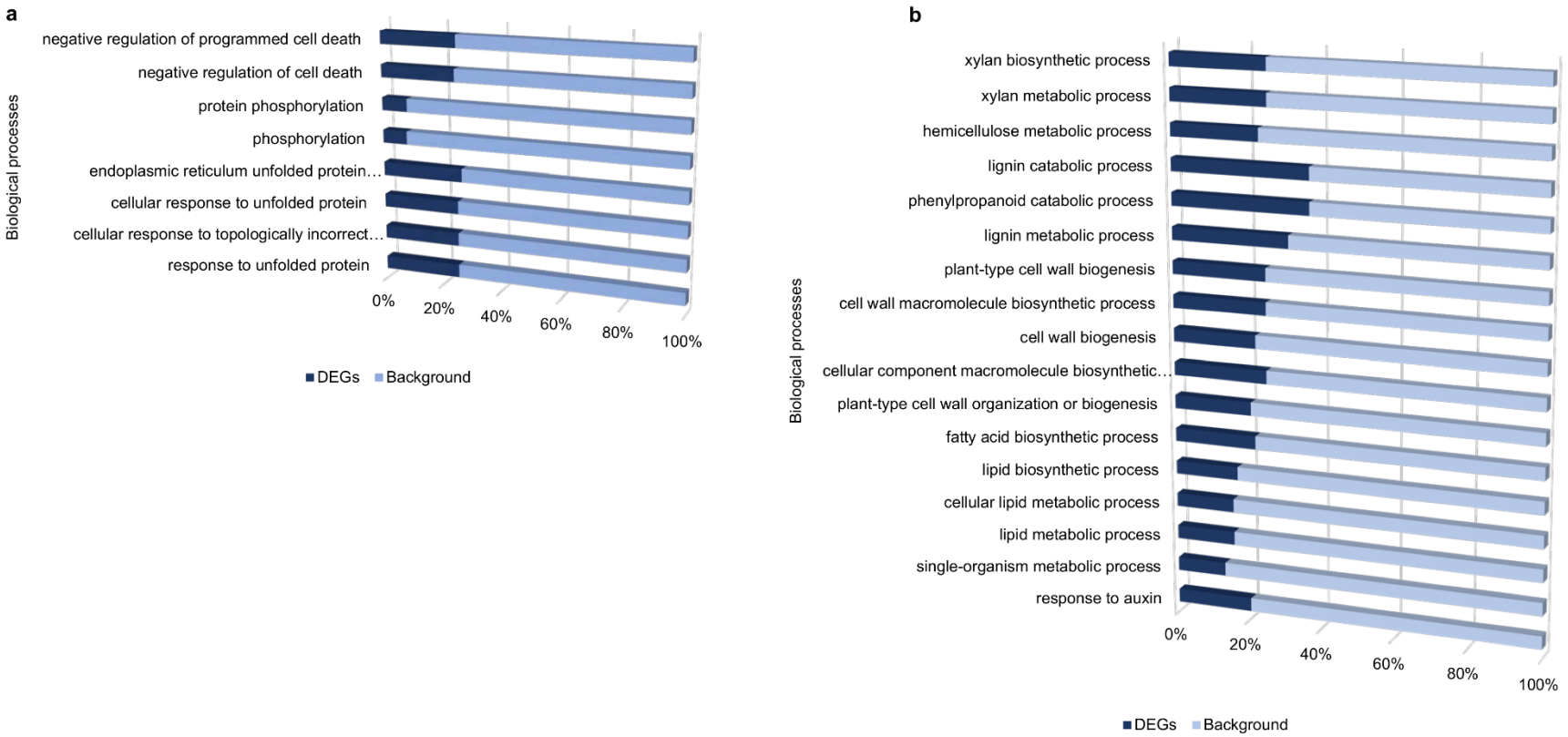


Fig. 2.3: Gene ontology enrichment analysis showing enriched processes specific to the tolerant cultivar (**a**), and specific to the susceptible cultivar (**b**), from DEGs identified when comparing inoculated samples to mock-inoculated controls.

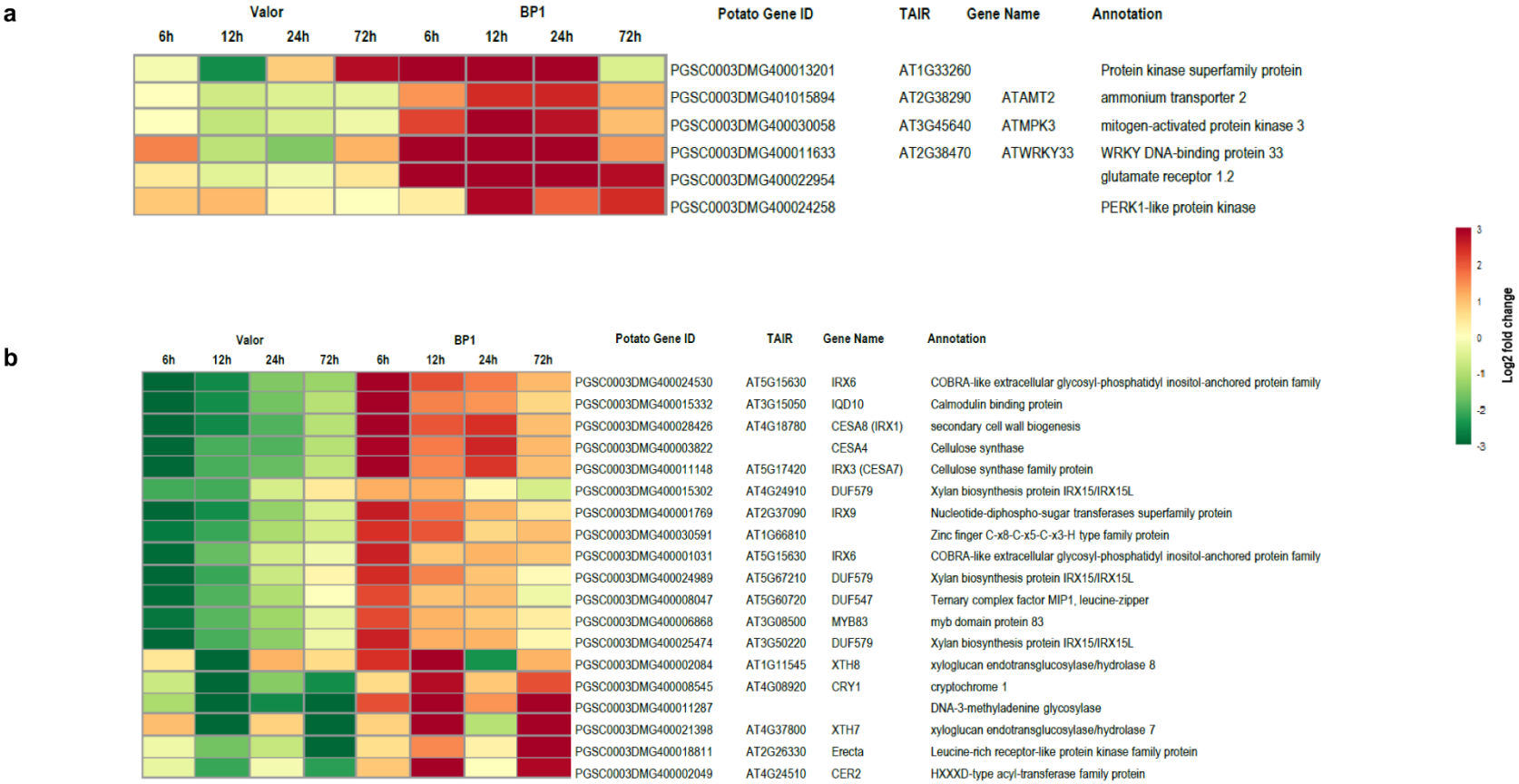


Fig. 2.4: Heat maps showing transcriptional profiles of selected DEGs enriched in cell communication and plant-type secondary cell wall biogenesis categories from the tolerant and susceptible cultivars, respectively. (**a**) DEGs important in plant defense responses, up-regulated in the tolerant cultivar. (**b**) Key secondary wall biosynthetic genes down-regulated in the susceptible cultivar.

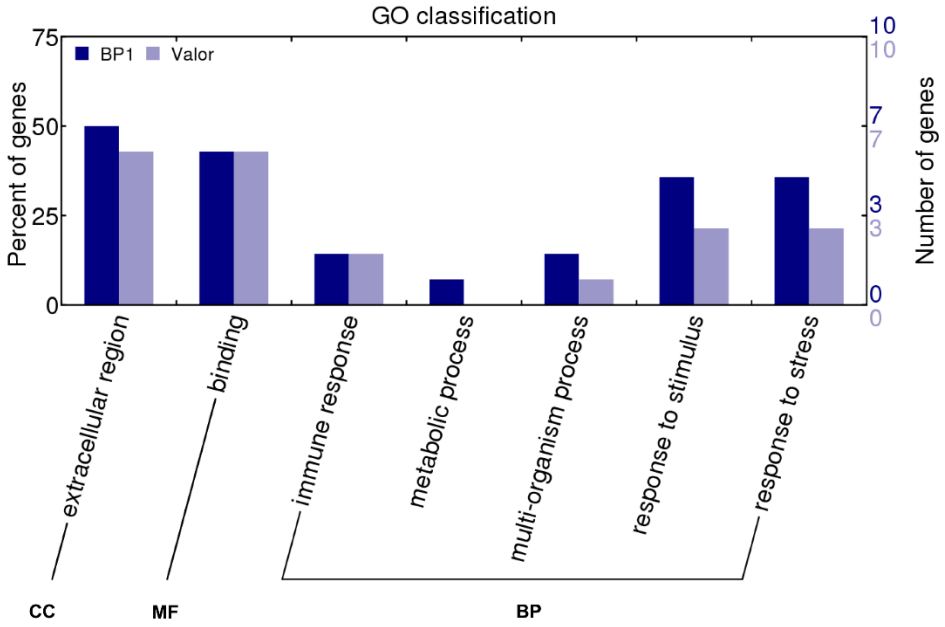


Fig. 2.5: GO classification of candidate novel CDS genes characterized using InterproScan5.

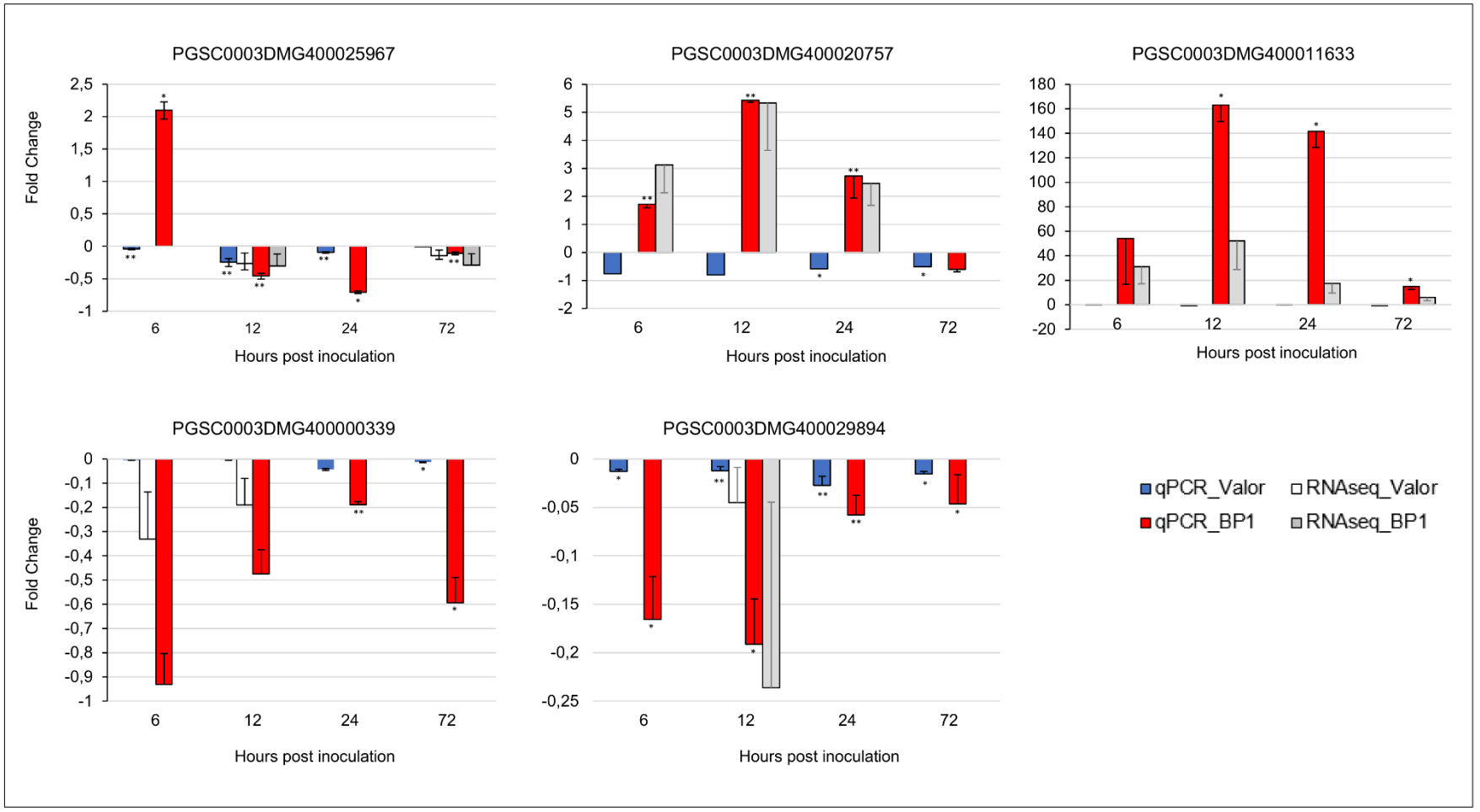


Fig. 2.6. RT-qPCR validation of RNA-seq gene expression ratios relative to mock inoculated samples using five selected DEGs. PGSC0003DMG400020757 (Membrane protein), PGSC0003DMG400029894 (Cytochrome P450 hydroxylase), PGSC0003DMG400025967 (Pectinesterase), PGSC0003DMG400000339 (Beta-galactosidase), PGSC0003DMG400011633 (AtWRKY33). Elongation factor 1-α (ef1α) and 18S RNA were used as the reference genes. Error bars represent the range of relative expression (qPCR fold change) calculated by $2^{-(\Delta\Delta Ct \pm SD)}$ (n=3). The RNA-seq bars at each time-point for each cultivar represent the fold changes calculated from three biological replicates using DESeq2 package, and the error bars represent log2 fold change standard error. Asterisks represent significant differences between inoculated samples and controls determined by Student's t-test (**P<0.01; *P<0.05).

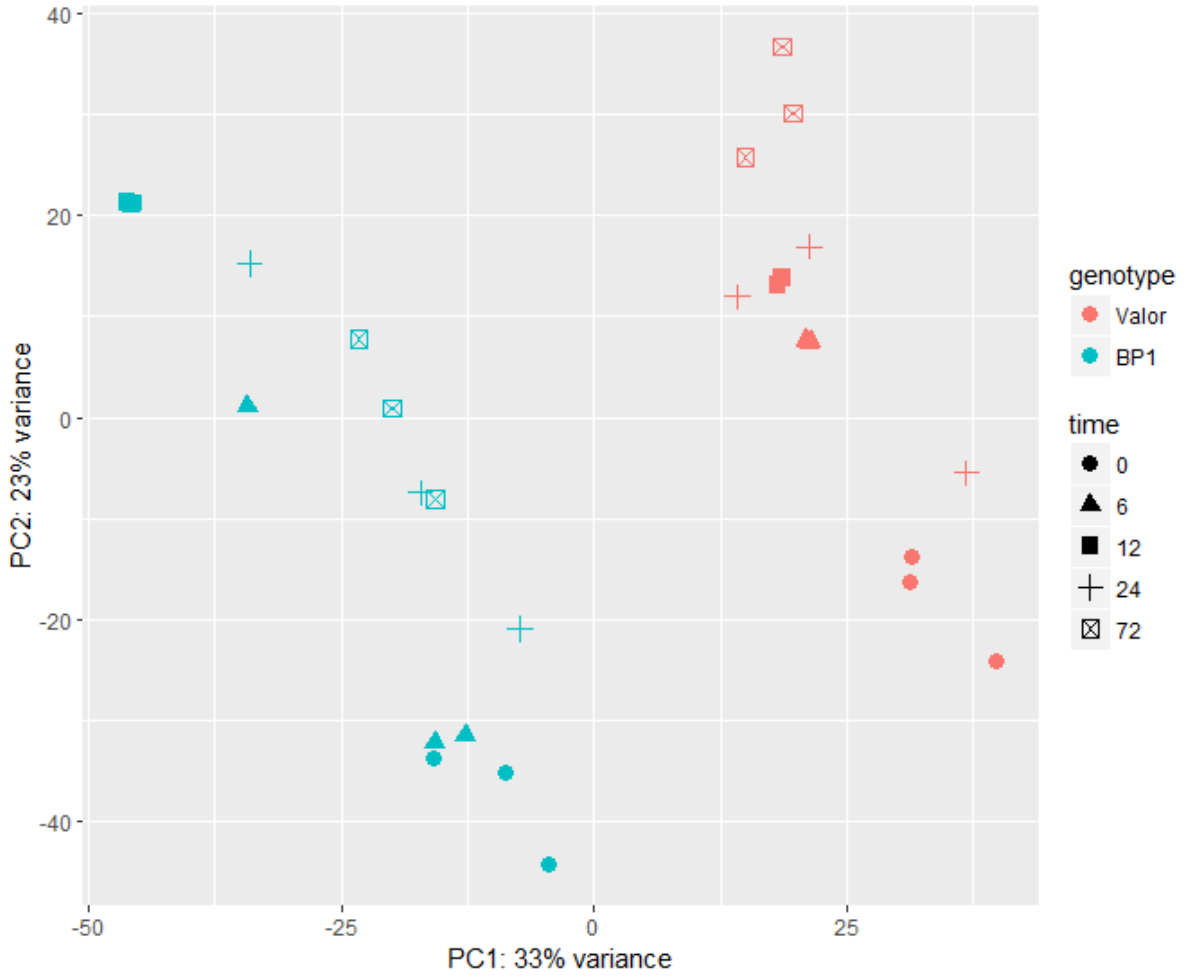


Fig. S2.1: Principal component analysis (PCA) plot showing congruence of RNA-seq data among biological replicates and indicating that the main differences in the dataset are due to the type of cultivar (i.e., differences between the susceptible and tolerant cultivars) and also influenced by time (i.e., different time-points used in this study). The PCA plot was generated using read count data transformed using the regularized-logarithm transformation (rlog) function implemented in DESeq2 package (Love et al., 2014).

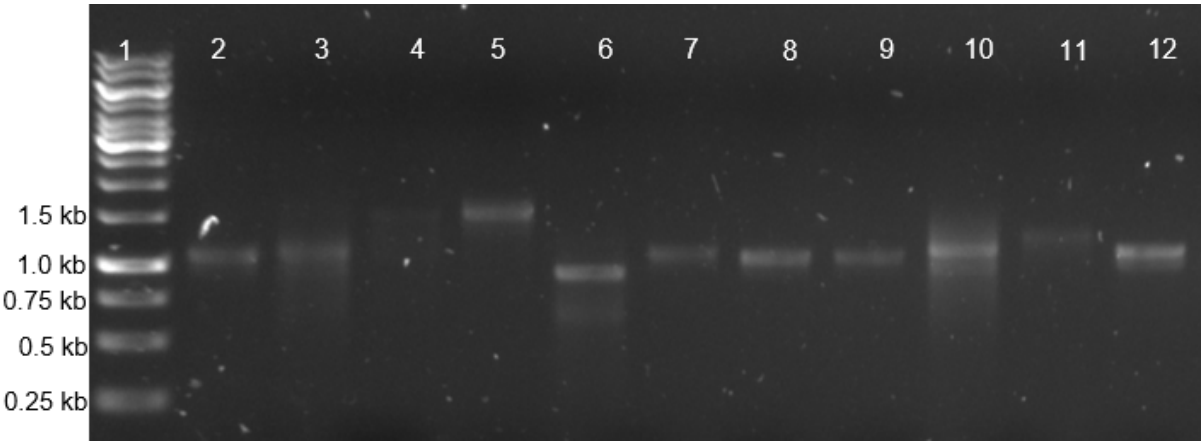


Fig. S2.2: RT-PCR validation of eight novel CDS candidates using agarose gel. Lane 1. 1 kb DNA ladder, Lane 2 and 3. Novel1253 in the tolerant and susceptible cultivars, respectively, Lane 4 and 5. Novel917 in the tolerant and susceptible cultivars, respectively, Lane 6, Novel806 in susceptible cultivar, Lane 7 and 9. Novel2142 in the susceptible cultivar, Lane 8. Novel1481 in the tolerant cultivar, Lane 10. Novel750 in the susceptible cultivar, Lane 11. Novel20149 in the susceptible cultivar, Lane 12. Novel2769 in the susceptible cultivar.

CHAPTER THREE

Genome-wide identification of potato long intergenic noncoding RNAs responsive to *Pectobacterium carotovorum* subspecies *brasiliense* infection

Stanford Kwenda[1], Paul R. J. Birch[2], Lucy N. Moleleki[1]

[1]Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0028, South Africa.

[2]The Division of Plant Sciences, College of Life Sciences, University of Dundee (The James Hutton Institute), Dundee DD25DA, Scotland, UK

This chapter has been prepared in the format of a manuscript and has been published in the peer-reviewed journal, BMC Genomics. I conceived the study together with Prof Lucy N. Moleleki. I performed the experimental work and bioinformatics analyses, and wrote the manuscript. Prof Lucy. N. Moleleki contributed in drafting the manuscript and obtained funding which supported this work. Prof Paul R. J. Birch, provided technical guidance and helped draft the manuscript.

3.1 Abstract

Background: Long noncoding RNAs (lncRNAs) represent a class of RNA molecules that are implicated in regulation of gene expression in both mammals and plants. While much progress has been made in determining the biological functions of lncRNAs in mammals, the functional roles of lncRNAs in plants are still poorly understood. Specifically, the roles of long intergenic nocoding RNAs (lincRNAs) in plant defence responses are yet to be fully explored.

Results: In this study, we used strand-specific RNA sequencing to identify 1113 lincRNAs in potato (*Solanum tuberosum*) from stem tissues. The lincRNAs are expressed from all 12 potato chromosomes and generally smaller in size compared to protein-coding genes. Like in other plants, most potato lincRNAs possess single exons. A time-course RNA-seq analysis between a tolerant and a susceptible potato cultivar showed that 559 lincRNAs are responsive to *Pectobacterium carotovorum* subsp. *brasiliense* challenge compared to mock-inoculated controls. Moreover, coexpression analysis revealed that 17 of these lincRNAs are highly associated with 12 potato defence-related genes.

Conclusions: Together, these results suggest that lincRNAs have potential functional roles in potato defence responses. Furthermore, this work provides the first library of potato lincRNAs and a set of novel lincRNAs implicated in potato defences against *P. carotovorum* subsp. *brasiliense*, a member of the soft rot *Enterobacteriaceae* phytopathogens.

3.2 Introduction

Advances in transcriptome profiling techniques especially with the advent of deep sequencing approaches (RNA-sequencing) have revealed that transcription in eukaryotes is much more complex than previously anticipated. It is now apparent that the bulk of eukaryotic genomes is pervasively transcribed giving rise to noncoding RNAs (ncRNAs) which exert pivotal effects on gene regulation (Bai et al., 2014). Noncoding RNAs can be grouped based on their lengths, into either (1) short ncRNAs (<200 bp) which have been extensively studied and generally include microRNAs (miRNAs), small nucleolar RNAs (snoRNAs), small nuclear RNA (snRNAs), and small interfering RNAs (siRNAs); and (2) long ncRNAs which are generally greater than 200 bp in length. Like mRNAs, lncRNAs have a 5' cap and a 3' poly-A tail; are mostly localized within the nucleus (Wierzbicki, 2012, Zhang and Chen, 2013), and can be multi-exonic (Ma et al., 2013). LncRNAs can exhibit cell or tissue specific expression patterns and have been observed to show poor conservation across different species (Ma et al., 2013). Based on their genomic location and context, lncRNAs are classified into intergenic (long intergenic noncoding RNA; lincRNA), long intronic noncoding RNA, and natural antisense transcripts (NATs). Natural antisense transcripts are RNA molecules with complementarity to other transcripts and can be grouped into *cis*-NATs (NATs fully antisense to protein coding genes on opposite strand) and *trans*-NATs (NATs with partial complementarity and transcribed from different loci) (Bai et al., 2014). Some lincRNAs can be located in close proximity to protein-coding genes (CDS), thus, may be referred to as adjacent-lncRNAs, and usually associated with CDS promoter and terminator regions. Furthermore, lincRNAs on one strand can partially overlap with CDS regions on the opposite strand and such lincRNAs may be termed antisense-lncRNAs.

In the past decade, much progress has been made towards understanding the roles of small non coding RNAs in plants (Liu et al., 2015). However, unlike small RNAs, the regulatory roles of lncRNAs remain poorly understood. Furthermore, compared to human and animal species, genome-wide discovery of lncRNAs in plants is still in its infancy (Kim and Sung, 2012).

Consequently, lncRNAs in plants constitute a class of ncRNAs that is less well-characterized. Nonetheless, regulatory roles of plant lncRNAs are now beginning to be recognized in diverse plant species through employing whole genome tilling arrays, *in silico* predictions and RNA-seq approaches (Liu et al., 2012a, Xin et al., 2011, Zhu et al., 2014, Li et al., 2014, Gallart et al., 2016). These emerging evidences demonstrate that lncRNAs play important roles in diverse biological processes in plants ranging from plant reproductive development, and responses to biotic and abiotic stresses (Zhu et al., 2014, Zhang et al., 2014, Amor et al., 2009).

The functional mechanisms of lncRNAs in many plant species are not yet fully understood with only a few lncRNAs having been fully characterized. In Arabidopsis, lncRNAs such as COLDAIR (cold-assisted intronic non-coding RNA) and COOLAIR (cold induced long antisense intragenic RNA) have been demonstrated to mediate chromatin modifying activities in transcriptional silencing of *FLC* during vernalization (Swiezewski et al., 2009, Heo and Sung, 2011). Another antisense lncRNA, ASL, a non-polyadenylated transcript, was recently discovered, and is implicated in epigenetic silencing of *FLC* (Shin and Chekanova, 2014). Additional regulatory functions of some lincRNAs such as *AT4* and *IPS1* (*INDUCED BY PHOSPHATE STARVATION1*) involve acting as decoys of miRNAs by a target mimicry mechanism, thus sequestering the regulatory roles of miRNAs away from their intended target genes (Wu et al., 2013, Franco-Zorrilla et al., 2007, Shin et al., 2006). It has recently been suggested that the Alternative Splicing Competitor long noncoding RNA (*ASCO*-lncRNA) also acts as a decoy, regulating gene expression in Arabidopsis during development (Bardou et al., 2014). The *ASCO*-lncRNA acts by competing to bind alternative splicing (AS) regulators, thus, diverting them from their AS mRNA targets (Bardou et al., 2014). Furthermore, plant lincRNAs have been implicated in important biological roles in responses to external stimuli (Franco-Zorrilla et al., 2007, Liu et al., 2012b, Shin et al., 2006, Shuai et al., 2014). In plants, genome-wide analysis of lncRNAs using deep sequencing transcriptomic data (mainly from RNA-seq approaches) have been performed on only a few plant species including Arabidopsis

(Amor et al., 2009, Zhu et al., 2014, Song et al., 2009), *Triticum* aestivum (Xin et al., 2011), *Medicago truncatula* (Wen et al., 2007), *Oryza sativa* (Zhang et al., 2014) tomato (Zhu et al., 2015) and *Zea mays* (Boerner and McGinnis, 2012, Li et al., 2014). Recently, a computational genome scale investigation of lncRNAs associated with annotated gene models was performed on 37 plant species including identification of 6788 potato (*Solanum tuberosum*) lncRNAs (Gallart et al., 2016). However, to date, investigation of the pervasive transcription in intergenic regions in potato and identification of lincRNAs have not yet been done on a genome-wide scale.

Potato is an important staple crop ranking fourth in global production after maize, rice and wheat. It can be severely affected by soft rot *Enterobacteriaceae* (SRE) species, in particular, *Pectobacterium carotovorum* subsp *brasiliense* (*Pcb*), an emerging member of the SRE, which is the most important causal agent of potato blackleg and soft rot globally including South Africa. Consequently, this pathogen poses a major threat to the potato industry in terms of yield, tuber quality and tuber seed exports (van der Merwe et al., 2010). Pathogen-responsive lincRNAs have been implicated in defence responses against *Fusarium oxysporum* infection in Arabidopsis (Zhu et al., 2014), and powdery mildew infection responses in wheat (Xin et al., 2011). Given the importance, albeit not well characterized, of lincRNAs in plant response to these pathogens, it would be interesting to unravel the repertoire of lincRNAs in potato and identify those responsive to this important emerging soft rot bacterium.

We identified 1113 potato candidate lincRNAs present in two potato cultivars that are susceptible (*S. tuberosum* cv. Valor) and tolerant (*S. tuberosum* cv. BP1) to *Pcb*. Using potato time-course RNA-seq data following infection with *Pcb*, we identified 559 potato lincRNA candidates that showed significant differential expression in the stems of the resistant and susceptible cultivars, compared to the mock-inoculated samples. Of these, six were validated using RT-qPCR. Importantly, expression of 17 lincRNAs was highly correlated with potato defence-related genes. Thus, our results suggest that lincRNAs are involved in potato defence mechanisms.

3.3 Results

3.3.1 Genome-wide identification of lincRNAs in potato

In order to identify long intergenic noncoding RNA (lincRNAs) related to potato defence networks, we employed a computational approach using strand-specific RNA-seq (ssRNA-seq) data derived from stems of *Solanum tuberosum* cultivars Valor and BP1 (Fig. 3.1). Samples of potato stems of each cultivar were harvested from six time-points post inoculation with *Pcb1692*, RNA isolated and pooled together. Sequencing was conducted on representative RNA pools of the susceptible and tolerant cultivars. The ssRNA-seq generated approximately 36 million (33 million uniquely mapped) and 38 million (35.3 million uniquely mapped) paired-end reads in *S. tuberosum* cvs Valor and BP1, respectively. From these data, a computational strategy was used that enabled the identification of lincRNAs after read mapping and transcript abundance assembly using Tophat2 (v2.0.13) and Cufflinks (v2.2.1), respectively (Trapnell et al., 2012, Trapnell et al., 2009). As an initial step, all transcript loci, from the potato genome annotation without strand information were removed prior to performing read alignments and transcript assembly. Subsequently, 59,681 and 60,292 transcripts were reconstructed for *S. tuberosum* cvs Valor and BP1, respectively. The majority of the assembled transcripts (84.7%) represented annotated genes and allelic isoforms in the potato reference genome assembly (PGSC_DM_v4.03) for both *S. tuberosum* cvs Valor and BP1.

3.3.2 Identification of novel transcriptionally active regions

To identify transcripts representing novel transcriptionally active regions (TARs) from the Cufflinks assembled transcripts, we first eliminated all the transcripts that overlapped with annotated potato features on the same strand. Our main focus in this study was particularly to identify novel ncRNA transcription; thus, only transcripts at a distance of more than 200 nucleotides from known genes on the same strand were considered, with lengths above 200 bp. Furthermore, to eliminate the possibility of genomic DNA contamination, only transcripts

with a sequencing depth of at least two reads per transcript were retained (Fig. 3.1). Additionally, since we were interested in measuring and comparing the variation in transcript abundances of the TARs between *S. tuberosum* cvs Valor and BP1 using time-course RNA-seq data in our downstream analysis, we started by first determining novel transcripts that were common between both the cultivars based on strand-specific RNA-seq data. Using the IntersectBed tool (v2.22.1) (Quinlan and Hall, 2010) and Blastn (Evalue: 1.0E-100) we identified 2950 novel transcripts that were present in the two cultivars. To determine a set of long intergenic noncoding RNA transcripts (lincRNAs) that is novel, the coding capability of these transcripts was then assessed using the Coding Potential Calculator (CPC) (Kong et al., 2007). CPC evaluates the protein-coding potential of transcripts based on prediction and assessment of potential open reading frames (ORFs) features and BLASTX (E-value cut-off 1.0E-10) homology searches against the non-redundant Uniprot Reference Clusters (UniRef90) protein database. Based on the extracted feature information, CPC algorithm, uses a score to classify transcripts into either protein-coding or noncoding. In this regard, all the transcripts showing evidence for protein-coding (CPC score > 0) were eliminated. Consequently, we obtained 1654 lincRNAs expressed in both potato cultivars with CPC scores less than zero. The 1654 lincRNAs obtained were further filtered to remove any lincRNAs with similarity to potato ribosomal DNA sequences (obtained from EnsemblPlants SolTub_3.0 Assembly (Blastn: E-value 1.0E-2)) resulting in 1649 lincRNA candidates. Of these, 1177 were high-confidence novel lincRNAs (CPC score < -1) and 472 were weak-novel lincRNA based on the CPC scores (Supplementary Table S3.1). However, because CPC uses a stringent Blastx cutoff (E-value: 1.0E-10), and only performs similarity analysis against the UniRef90 protein database, it is possible that some mRNA transcripts with relatively weak protein signatures could be falsely classified as potential lincRNA transcripts. Thus, we further screened the 1649 lincRNA candidates against InterPro (Mitchell et al., 2014), using InterProScan5 (Jones et al., 2014). LincRNA sequences with similarities to protein families and domains from any of the databases within the InterPro consortium were considered as protein-coding and eliminated. Finally, following the additional filters, a list of 1113 transcripts

was regarded as novel lincRNA potato transcripts expressed in stems of *S. tuberosum* cvs Valor and BP1 (Supplementary Table S3.1). Semi-quantitative reverse transcription (RT)-PCR confirmed nine of the RNA-seq identified lincRNAs, thus validating the assembly quality and identification pipeline (Fig. 3.2).

3.3.3 Characterization and classification of potato lincRNAs

Using basic features of the identified lincRNAs in a genomic context, we found that the lincRNAs ranged from 200 to 17,256 bp in size and were transcribed from all the 12 potato chromosomes (Potato Genome Assembly: PGSC_DM_v4.03) (Supplementary Fig. S3.1a). The highest and least numbers of lincRNAs were transcribed from chromosome one (183 lincRNAs) and chromosome 12 (nine lincRNAs), respectively. As with mRNA transcripts, lincRNAs appeared to be distributed uniformly across all chromosomes, with the exception of chromosome 12, were lincRNAs were only concentrated within the region up to 5 Mbp (Supplementary Fig. S3.1a). In addition, based on length distribution, lincRNAs can be divided into three groups, namely, short-length, medium-length and long-length lincRNAs (Ma et al., 2013). Thus, the majority of potato lincRNAs (71%) are short-length lincRNA (200-1000 bp), 26% are medium-length lincRNA (1-5 kb) and only 3% are long-length lincRNA (> 5 kb). In contrast, most of the protein-coding transcripts, 54% comprise medium transcripts (Supplementary Fig. S3.1b). Comparing the number of exons between annotated potato genes and lincRNAs showed that on average, lincRNAs possess fewer exons (Table 3.1). Furthermore, we assessed the repeat content (including presence of transposons) of potato lincRNAs using RepeatMasker (http://www.repeatmasker.org) and the TIGR *Solanum* Repeat Database v3.2 (plantrepeats.plantbiology.msu.edu/downloads.html). Almost half of the lincRNAs (42.3%) contain repetitive sequences.

Even though it is still not yet clear how classification of lncRNAs based on their proximity to coding genes reflects biological function, knowing where lncRNAs are located in the genome and their expression profiles provides useful insights into their biological significance and

primary mechanisms of action (Atkinson et al., 2012). Thus, we classified the identified lincRNAs into three types, based on their genomic location and proximity with respect to their closest protein-coding genes, namely: intergenic (distance > 1kb; without any overlaps with CDSs on both strands), adjacent (distance < 1kb) and antisense-lncRNAs (those partially overlapping protein-coding genes on the opposite strand) (Fig. 3.3a). Most of the lincRNAs (87%) are located at least 1kb away from annotated potato gene models on either strand, 8% of lincRNA are adjacent-lncRNA, located in close proximity to protein-coding genes, and only a small proportion (5%) constituted antisense-lncRNAs (Fig. 3.3b and Supplementary Table S3.1). The percentage difference observed between lincRNAs at distance > 1kb from CDS regions and antisense-lncRNA is consistent with previous observations made in maize (Li et al., 2014) and tomato (Zhu et al., 2015).

Furthermore, the identified lincRNA sequences were compared with lncRNA sequences from tomato (Zhu et al., 2015), Populus (Shuai et al., 2014), and Arabidopsis (Jin et al., 2013, Yi et al., 2015, Liu et al., 2012b) to determine the set of potato lincRNAs with similarity to these plant species (BLASTn e-value < 1.0E-10). As expected, the lincRNAs displayed poor conservation. Only 13% lincRNAs showed multiple homologous regions (>80% identity and alignment length > 100 bp) with 231 lncRNAs from tomato and two from Arabidopsis (Supplementary Table S3.2). Thus, unlike most mRNAs which are highly conserved across organisms, lincRNAs tend to evolve rapidly resulting in poor conservation (Ulitsky and Bartel, 2013). Lastly, to check the novelty of our set of lincRNAs, we checked for overlaps against potato lncRNAs reported by Gallart *et al*. (Gallart et al., 2016). Comparisons using IntersectBed tool (v2.22.1), showed that only nine lincRNAs out of the 1113 lincRNAs from our set overlapped with and were similar to nine previously reported potato lncRNAs (Supplementary Fig. S3.2).

3.3.4 Quantitative analysis of potato lncRNAs responsive to *Pectobacterium carotovorum* subspecies *brasiliense* infection

From our previous work (Kubheka et al., 2013), we showed that *S. tuberosum* cv. Valor is highly susceptible to *Pcb1692* infection showing typical blackleg symptoms upon infection. On the other hand, *S. tuberosum* cv. BP1 was shown to be tolerant to *Pcb1692*. Furthermore, gene expression analysis performed in Chapter 2 between cultivars 'Valor' and 'BP1' revealed differentially expressed protein-coding genes involved in potato defence responses to *Pcb1692* infection. However, it remains to be investigated whether lincRNA expression is activated in response to *Pcb1692* infection in potato. Thus, we hypothesized that lincRNAs could be involved in potato defence mechanisms and therefore differentially expressed in the tolerant compared to susceptible cultivar. Consequently, in the present work, we sought to determine novel lincRNA transcripts that were differentially expressed between the two cultivars and could thus be implicated in potato defences against the necrotrophic plant pathogen, *Pcb1692*. Our bioinformatics analysis showed that a total of 1113 lincRNAs were expressed in both cultivars. Thus, to identify defence-related lincRNAs, the expression levels of these commonly expressed 1113 lincRNAs were compared between the two cultivars, *S. tuberosum* cvs Valor and BP1. To do this, each cultivar was inoculated with *Pcb1692* (1×10^9 $cfu.ml^{-1}$) and samples obtained at five different time points (0 (Buffer inoculated control), 6, 12, 24, 72 hours post inoculation; hpi). RNA was isolated from each time point and three biological replications per time point were prepared and sequenced independently. Using the resulting time-course RNA-seq data, the differential expression patterns of the 1113 lincRNAs was evaluated. In total, 485, 416, 539, 364, and 449, lincRNAs were differentially expressed (DE) at 0, 6, 12, 24, and 72 hpi, respectively, between *S. tuberosum* cvs *Valor* and *BP1*, at 10% false discovery rate (FDR) (Supplementary Table S3.3). Compared with *S. tuberosum* cv *Valor*, an average of 51% lincRNAs were upregulated in *S. tuberosum* cv *BP1*, throughout the time-course. Numbers of up-regulated lincRNAs in the tolerant cultivar were slightly higher at 6, 12 and 72 hpi, with the most up-regulated lincRNAs observed at 6 hpi (54%) (Fig 3.4a).

Furthermore, in order to determine the expression profiles of these differentially expressed lincRNAs showing cultivar-specific differences, DE lincRNAs at each time-point (post inoculation) were compared to mock-inoculated samples (0 hpi) in each cultivar. In total, 173 DE lincRNAs were present in both cultivars and 267 and 119 were only significantly expressed in *S. tuberosum* cv *Valor* and *S. tuberosum* cv *BP1*, respectively (Fig. 3.4b and 3.4c).

To confirm RNA-seq expression patterns and determine whether the differentially expressed lincRNAs are involved in potato defence responses, six of these lincRNAs were arbitrarily selected representing lincRNAs that were up or down regulated at one or more time-points and validated experimentally using reverse-transcription quantitative PCR (RT-qPCR) (Fig. 3.5). The RT-qPCR results were in concordance with the RNA-seq data, thus, implicating the DE lincRNAs in potato defence responses.

3.3.5 LincRNA/ mRNA genes expression correlation

In order to understand the possible biological roles of the differentially expressed (DE) lincRNAs in relation to potato defence responses, we investigated all-against-all coexpression patterns between lincRNA transcripts and DE mRNA genes within the time-course using hierarchical clustering. In total, 179 lincRNAs were highly correlated with 3,573 mRNA genes (Spearman rank correlation (r_{rho}) > |0.8| and were clustered into 62 different groups. Interestingly, Gene Ontology (GO) enrichment analysis using the Panther Classification System web server (Mi et al., 2016), showed that 32 clusters contained CDS genes enriched in “response to stimulus”, including secondary GO terms such as “defence response to bacterium”, “response to stress” and “response to endogenous stimulus”. Therefore, to highlight potential lincRNA functions and/ or interactions with CDS genes involved in potato defence mechanisms, we further performed pairwise correlations between CDS genes and lincRNAs within each cluster associated with response to stimulus GO terms. Overall, 17 lincRNAs exhibited extremely high positive correlation ($r_{rho} \geq 0.9$) with 12 potato defence-related CDS genes (Table 3.2). These results suggest that these highly correlated lincRNAs could be involved in potato defence responses against *Pcb1692* infection.

3.3.6 Prediction of interactions between lincRNAs and miRNAs

Long noncoding RNAs can be involved in diverse cellular molecular functions depending on their mode of action (Ma et al., 2013). Because lincRNAs are functional RNA molecules, they can be targeted and regulated by miRNAs post transcriptionally, triggering degradation of the targeted lincRNAs. To investigate whether the identified potato lincRNAs are targeted by miRNAs, we analyzed the 1113 lincRNAs using psRNATarget (Dai and Zhao, 2011). A total of 57 lincRNAs were predicted to be targeted by 98 potato miRNAs (Supplementary Table S3.4). Of these lincRNAs, four were targeted by six miRNAs implicated in plant immune defences (Li et al., 2012) (Supplementary Table S3.4). Interestingly, none of these four lincRNAs were differentially expressed in the time-course following inoculation with *Pcb1692* possibly reflecting their miRNA mediated cleavage and degradation. RT-qPCR analysis confirmed expression of these defence-related miRNAs under the same experimental conditions, adding credence to their possible interaction or regulation of their target lincRNAs (Supplementary Fig. S3.3). An additional 10 lincRNAs were targets of various members of stu-miR5303 family which is part of nine miRNA families unique to solanaceous plants (Gu et al., 2014) (Supplementary Table S3.4). Previously identified potential targets of stu-miRNA5303 family include proteins responsive to abiotic stress, metabolic enzymes and proteins of unknown function (Gu et al., 2014). The large numbers of stu-miRNA5303 members implies its biological importance. Consequently, it is plausible to assume that their target lincRNA transcripts play important biological roles in potato.

3.4 Discussion

The regulatory roles of lincRNAs are increasingly being unraveled in plants, as indicated by the number of various reports on the identification of lncRNAs in plant species including maize, rice, *Populus* and Arabidopsis (Shuai et al., 2014, Boerner and McGinnis, 2012, Liu et al., 2012a, Li et al., 2014, Song et al., 2009, Zhang et al., 2014, Liu et al., 2012b). However, most of these reports have focused on lncRNAs involved in plant development, reproduction and abiotic stress responses (Zhang and Chen, 2013, Amor et al., 2009, Wang et al., 2015, Li et al., 2014, Zhang et al., 2014). In contrast, reports about lincRNAs involved in defence regulatory mechanisms against pathogens are just beginning to emerge (Xin et al., 2011, Zhu et al., 2014). In potato, previous studies on ncRNA have predominantly focused on miRNA identification and functional analysis (Li et al., 2012, Xie et al., 2011, Zhang et al., 2013, Zhang et al., 2009), but no data have been reported for lincRNAs, especially in association with potato defence responses. In this study, we conducted a genome-wide analysis of potato lincRNAs, by integrating strand-specific RNA sequencing with time-course RNA-seq data. We identified novel candidate lincRNAs potentially associated with potato defence response mechanisms during challenge by *Pcb1692*. Hence, this present work provides an important resource of potato lincRNAs that can be useful to other researchers.

To facilitate the identification of lincRNAs, a strand-specific RNA-seq (ssRNA-seq) approach was employed which made it possible to determine the strand from which the lincRNAs were produced. This is in contrast to some previous lincRNA identification reports in plants, which had the limitation of RNA-seq data lacking strand information (Li et al., 2014, Shuai et al., 2014). Knowledge of the strand information of lncRNAs is important in localizing their genome context and position since lincRNAs are transcribed from intergenic regions with some being adjacent or antisense to protein-coding regions (Clark and Blackshaw, 2014). Thus, strand-specific RNA-seq allowed us to classify the identified lincRNAs into three categories, based on their proximity to protein-coding genes (Fig. 3.3b). Classification of lincRNAs based on their genomic location can be a useful preliminary step in determining potential functional roles of

lincRNAs (Ma et al., 2013). In addition, our present work revealed that, like protein-coding genes, lincRNAs are distributed throughout the potato genome (Supplementary Fig. S3.1a). Thus, the pervasive expression of lincRNAs in the entire 12 potato chromosomes suggests that they are common RNA molecules representing a functional component of the potato genome.

Additionally, by aligning the ssRNA-seq reads, using Tophat2, a splice-aware aligner and performing transcript reconstruction using the software tool Cufflinks (Trapnell et al., 2012), a parsimonious representation of exon boundaries for the lincRNAs was obtained. As a result, the structure of lincRNA transcripts was resolved. Furthermore, since consideration was only given to lincRNA candidates conserved in *S. tuberosum* cvs. Valor and BP1, the identified 1113 lincRNAs constitute a reliable list of lincRNAs from potato stems, extending the current understanding of the potato transcriptome landscape.

In general, functional characterization of lincRNAs in plants is still in its infancy. Moreover, little is known about regulatory functions of lincRNAs in biotic stress responses in plants. Currently, the function of lncRNAs cannot be inferred directly from primary sequence or structure as is the case with miRNAs and protein-coding mRNA (Mercer et al., 2009). However, key insights into biological roles of lncRNAs can be derived from the conditions in which they are expressed (Atkinson et al., 2012). In the present study, we identified 559 differentially expressed (DE) lincRNAs at different time-points (up to 72 hpi) in *S. tubersom* cvs. Valor and BP1, compared to mock-inoculated samples (Fig 3.4b). Meanwhile, the responsiveness of six DE lincRNAs to *Pcb1692* infection was also confirmed using RT-qPCR, further alluding to the potential functional activity of these lincRNAs.

From a systems biology perspective, the guilt-by-association principle has been applied successfully for the functional characterization of various genes in humans and other mammals assuming functional relationships between co-expressed genes (Saito et al., 2008, Wolfe et al., 2005). Thus, in the present study, a hierarchical clustering strategy was employed

in order to infer potential functional roles of DE lincRNAs in the time-course, and enriched functions of CDS genes within individual clusters were identified. Only clusters with genes enriched for biological process GO terms under the “response to stimulus” category were considered. Generally, coexpression analysis is used to predict biological processes and infer novel members (genes, ncRNA transcripts etc.) of known processes and/or pathways (Rhee and Mutwil, 2014). In addition, identifying lincRNAs associated with the “response to stimulus” category was particularly relevant in this study due to the fact that most defence-related genes are often overrepresented within this category. Therefore, GO terms of CDS genes that showed significantly high pairwise coexpression with lincRNAs ($r_{rho} > 0.9$) within the clusters, were mapped to lincRNAs (Table 3.2). Based on this analysis, 17 lincRNAs were co-expressed with genes associated with defence-related GO terms such as response to stimulus, defense to bacterium, response to endogenous stimulus, response to stress and response to toxic substance (Table 3.2). Most of these CDS genes belong to the Leucine-rich repeat protein kinase family, which mainly function as pattern recognition receptors in plant innate immune responses. Thus, our results implicate these co-expressed lincRNAs in defense responses against *Pcb1692* in the tolerant cultivar, making them key candidates for future experimental validations.

Interestingly, these 17 lincRNAs are located more than 100 kb from their correlated CDS genes, and some of them are interchromosomal with regards to CDS genes they are co-expressed with. Thus, these lincRNAs are possibly *trans*-acting, functioning as transcriptional regulators that interact with genes at distal locations across multiple chromosomes. However, there is a paucity of information regarding molecular mechanisms employed by lincRNAs in regulating their distal gene targets. Nonetheless, we anticipate that these mechanisms will become more apparent in the near future.

3.5 Conclusions

This study focused on the genome-wide discovery of lincRNAs using strand-specific RNA-seq and resulted in the first catalogue of potato lincRNAs, comprising 1113 transcripts, including 1104 novel lincRNA candidates, derived from stem tissue. In addition, we identified 559 lincRNAs that were responsive to *P. carotovorum* subsp. *brasiliense* infection in *S. tuberosum* cvs *Valor* and *BP1*. Importantly, 17 differentially expressed lincRNAs were highly associated with defence-related CDS genes, thus representing key candidates for future functional studies.

3.6 Materials and Methods

3.6.1 Plant material and growth conditions

Seed tubers of two potato cultivars, susceptible (*Solanum tuberosum* cv. Valor) and tolerant (*S. tuberosum* cv. BP1) to *Pectobacterium carotovorum* subsp *brasiliense* strain 1692 (*Pcb1692*) infection were grown in the greenhouse under standard conditions (22 to 26 °C, 16 h light/ 8 h dark photoperiod and 70% relative humidity). Stem inoculations were done as previously described in Kubheka *et al*. (Kubheka et al., 2013), except that we used wild-type *Pcb1692* for the inoculations and inoculated plants were assessed and sampled at 0, 6, 12, 24, and 72 hours post inoculation (hpi) in triplicates (three plants were pooled together for each biological replicate).

3.6.2 Total RNA preparation

Total RNA was extracted from potato stems using the QIAGEN RNeasy plant mini kit (Qiagen) including DNAse treatment (Qiagen). RNA was quantified using the NanoDrop (Thermo Scientific, Sugarland, TX, USA) and the quality and integrity checked using Agilent 2100 BioAnalyzer system (Agilent, Santa Clara, CA, USA).

3.6.3 Whole transcriptome library construction and sequencing

The construction of whole transcriptome libraries and sequencing were carried out at the Beijing Genomics Institute (BGI-Shenzhen, China). For the preparation of strand-specific libraries, total RNA was pooled from five time-points (0, 6, 12, 24, 72 hpi) for BP1 and Valor. Whole transcriptome libraries were constructed using the TruSeq Stranded RNA Sample Prep Kit v2 (Illumina, San Diego, CA), according to the manufacturer's instructions. For the time-course experiment, standard (normal) transcriptome libraries were constructed using RNA samples from individual biological replicates (n=3) from each time-point using the TruSeq RNA sample Prep Kit v2 (Illumina, San Diego, CA) following manufacturer's instructions. The libraries were quality checked and quantified using Agilent BioAnalyzer 2100 system and

qPCR. Finally, the libraries were sequenced on the Illumina HiSeq 2000 system generating 90 bp paired-end reads. The data have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through the GEO accession number, GSE74871.

3.6.4 Assembly of RNA transcripts

Strand-specific sequencing reads for each cultivar were quality checked using FASTQC (http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc) and mapped to the potato reference genome (Genome assembly: PGSC_DM_v4.03; http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml) using TopHat2 (version 2.0.13) [--library-type fr-firststrand –G] (Trapnell et al., 2009). For the alignments, the minimum (-i) and maximum (-I) intron sizes were obtained at http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml, and set at 10 bp and 15000bp, respectively. Transcript assembly was performed using Cufflinks (version 2.2.1) [-g -u --library-type fr-firststrand] (Trapnell et al., 2012).

3.6.5 Bioinformatics identification of lincRNAs

The assembled potato transcripts were compared with annotated potato protein sequences (http://potato.plantbiology.msu.edu/data/PGSC_DM_V403_representative_genes.gff.zip) using IntersectBed (v2.22.1) (Quinlan and Hall, 2010). All assembled transcripts overlapping with potato coding sequences and less than 200 bp from protein coding regions were removed. For size selection, java scripts were used to filter out all transcripts less than 200 nucleotides in length. For the sequencing depth filter, HTSeq-count [python -m HTSeq.scripts.count -f bam -s reverse] (Anders et al., 2014) was used and only transcripts with at least two reads were considered. Following sequencing depth filter, IntersectBed (v2.22.1) (Quinlan and Hall, 2010) and stringent Blastn (Evalue: 1.0E-100) was used to extract novel transcripts present in both BP1 and Valor. Since lncRNA transcripts are generally known not to have any coding capacity, all the transcripts common to BP1 and Valor were tested for protein-coding potential using the Coding Potential Calculator (CPC) (Kong et al., 2007).

Following the coding potential filter, only transcripts with a negative CPC score were retained as potential novel lincRNA candidates.

3.6.6 Distribution of lincRNAs and protein-coding genes in the potato genome

A circular representation of the distribution of lincRNAs and mRNAs was constructed using Circos (Krzywinski et al., 2009) for comparative visualizations among the 12 chromosomes.

3.6.7 Classification of lincRNAs

Potato lincRNAs were classified into three categories based on their genomic location and distance from protein-coding genes nearest to each lincRNA transcript using IntersectBed (v2.22.1) (Quinlan and Hall, 2010) and java scripts. The lincRNAs were grouped into: 1) intergenic-lncRNA, without any overlaps with protein-coding genes on both strands and at least 1kb away from the nearest CDS 2) adjacent-lncRNA, which are in close proximity to protein coding genes but without any overlaps and 3) antisense-lncRNA, which partially overlap with genes on the opposite strand.

3.6.8 Differential expression analysis of lincRNAs between the tolerant and susceptible potato cultivars

Time course RNA-seq data from stems of BP1 and Valor was used to identify lincRNAs responsive to *P. carotovorum* subsp. *brasiliense* infection. Briefly, to identify differentially expressed lincRNAs between Valor and BP1, RNA-seq reads were quality checked using FASTQC and mapped to the potato reference genome using TopHat2 (Trapnell et al., 2009). HTSeq-count was used to make read counts mapped to lincRNA transcripts and DeSeq2 (Love et al., 2014) was used to determine the differential expression with a false discovery rate threshold of 10%.

3.6.9 Quantitative reverse transcription PCR (RT-qPCR)

For RT-qPCR, first-strand cDNA synthesis was done from total RNA using Superscript III First-Strand cDNA Synthesis SuperMix kit (Invitrogen, USA) following manufacturer's instructions. Quantitative real-time PCR using Applied Biosystems SYBR Green Master Mix was performed in the QuantStudio 12K Flex Real-Time PCR system (Life Technologies, Carlsbad, CA, USA). For RT-qPCR, 2 µl of sample was added to 8 µl of Applied Biosystems SYBR Green Master Mix and primers at a concentration of 0.4 µM. The cycling conditions were as follows: an initial denaturation at 50 °C for 5 min and 95 °C for 2 min followed by 45 cycles of 95 °C for 15 s and 60 °C for 1 min. Each sample was run in triplicate and two biological replicates were employed. The samples were normalized to 18S rRNA and elongation factor 1-*α* (ef1*α*) as the reference genes (Nicot et al., 2005) and the mock treated samples used as calibrators. The comparative CT ($\Delta\Delta^{ct}$) method was used to measure relative expression (Livak and Schmittgen, 2001). Primers used were designed online using Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and are listed in Supplementary Table S3.5.

3.6.10 LincRNA-mRNA coexpression analysis

To investigate correlations of expression between lincRNA and differentially expressed mRNA transcripts under the same conditions, an all-against-all hierarchical clustering analysis was performed based on log2 fold changes using Cluster 3.0 software (Eisen et al., 1998). Briefly, LincRNA and mRNA datasets were filtered so that only transcripts with an expression of at least 2-fold at any of the 5 time-points tested were considered. Clustering was performed using the Spearman Rank Correlation similarity metric ($r_{rho} > |0.8|$) and the complete linkage clustering method. Visualization was done using TreeView program (Saldanha, 2004). In order to predict potential lincRNA functions, mRNA transcripts grouped together with lincRNAs in various clusters were used to perform Gene Ontology (GO) analysis based on the Panther Classification System (version 10.0) web server (Mi et al., 2016). Corresponding orthologs in Arabidopsis of the differentially expressed potato mRNA genes were used for the GO

enrichment analysis, based on BLASTp (e-value: 1.0E-05), implemented in ProteinOrtho program (Lechner et al., 2011). Lastly, pairwise Spearman correlation coefficient was calculated by cor.test() in R, between CDS genes and lincRNAs grouped within each cluster associated with response to stimulus GO biological process terms. To assign putative functional annotations to the lincRNAs, GO terms of CDS genes significantly correlated with lincRNAs were mapped to the lincRNAs.

3.6.11 RT-PCR validation of lincRNA transcripts

First-strand cDNA was synthesized as described above for RT-qPCR and the PCR was performed on Bio-RAD T100TM Thermal Cycler conventional PCR (Bio-RAD, USA). The lincRNA primers were designed online using Primer3plus (Supplementary Table S3.6). PCR was performed in a 25 µl reaction mix containing 1 µl of template cDNA (~40 ng), Taq DNA Polymerase, 10x Taq Buffer (New England Biolabs, UK), 2.5 mM dNTPs each and 0. 5µM of forward and reverse primer each. Thermal cycling conditions were: 95 °C for 2 min; 30 cycles of 95 °C for 30 sec, 57 °C for 30 sec, 72 °C for 60 sec, and the final extension at 72°C for 5 min. The PCR products were analysed on 1.5% agarose gel including 1 kb DNA molecular weight ladder 470 (NEB, UK). To check for genomic DNA contamination, a non reverse-transcriptase control was included.

3.6.12 Prediction of lincRNA and miRNA interactions

Potato lincRNAs targeted by miRNAs were predicted using the psRNATarget (Dai and Zhao, 2011) server by using default parameters.

3.7 Supplementary Data

Fig. S3.1: (**a**) Comparison of the genomic distribution of lincRNAs and protein-coding genes across the 12 potato chromosomes. The outer grey track represents the 12 potato chromosomes, with a scale (Mb) showing the length of each chromosome. The red histograms (second track with an outer orientation) and blue histograms (third track with inner orientation) represent the abundance and distribution of mRNA and lincRNAs, respectively, throughout the potato genome. The bin size (histogram width) = 5 Mbp). (**b**) Comparison of LincRNA lengths to protein-coding mRNA transcripts in potato (PGSC_DM_v4.03 genome assembly).

Fig. S3.2: Comparison of the 1113 lincRNA transcripts identified in the present study with potato lncRNAs available in the GreenC database.

Fig. S3.3: RT-qPCR confirmation of five potato defense-related miRNAs in *S. tuberosum* cv BP1, computationally predicted to target some of the lincRNA transcripts. U6 snRNA was used as the reference gene. The fold changes of miRNAs at each time point were calculated relative to calibrator (control sample; 0 hpi). The experiments were done in triplicate. Error bars represent the fold change range calculated by 2-(ΔΔCt±SD).

Table S3.1: List of the identified 1113 lincRNA candidates

Table S3.2: LincRNA conservation analysis

Table S3.3: Differentially expressed lincRNA transcripts

Table S3.4: LincRNAs targeted by potato miRNAs

Table S3.5: List of RT-qPCR primers used in this study

3.8 References

AMOR, B. B., WIRTH, S., MERCHAN, F., LAPORTE, P., D'AUBENTON-CARAFA, Y., HIRSCH, J., MAIZEL, A., MALLORY, A., LUCAS, A. & DERAGON, J. M. 2009. Novel long non-protein coding RNAs involved in Arabidopsis differentiation and stress responses. *Genome Research,* 19**,** 57-69.

ANDERS, S., PYL, P. T. & HUBER, W. 2014. HTSeq–A Python framework to work with high-throughput sequencing data. *Bioinformatics***,** btu638.

ATKINSON, S. R., MARGUERAT, S. & BÄHLER, J. Exploring long non-coding RNAs through sequencing. Seminars in cell & developmental biology, 2012. Elsevier, 200-205.

BAI, Y., DAI, X., HARRISON, A. P. & CHEN, M. 2014. RNA regulatory networks in animals and plants: a long noncoding RNA perspective. *Briefings in Functional Genomics***,** elu017.

BARDOU, F., ARIEL, F., SIMPSON, C. G., ROMERO-BARRIOS, N., LAPORTE, P., BALZERGUE, S., BROWN, J. W. & CRESPI, M. 2014. Long noncoding RNA modulates alternative splicing regulators in Arabidopsis. *Developmental Cell,* 30**,** 166-176.

BOERNER, S. & MCGINNIS, K. M. 2012. Computational identification and functional predictions of long noncoding RNA in Zea mays. *PLoS One,* 7**,** e43047.

CLARK, B. S. & BLACKSHAW, S. 2014. Long non-coding RNA-dependent transcriptional regulation in neuronal development and disease. *Frontiers in Genetics,* 5.

DAI, X. & ZHAO, P. X. 2011. psRNATarget: a plant small RNA target analysis server. *Nucleic Acids Research,* 39**,** W155-W159.

EISEN, M. B., SPELLMAN, P. T., BROWN, P. O. & BOTSTEIN, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Science USA,* 95.

FRANCO-ZORRILLA, J. M., VALLI, A., TODESCO, M., MATEOS, I., PUGA, M. I., RUBIO-SOMOZA, I., LEYVA, A., WEIGEL, D., GARCÍA, J. A. & PAZ-ARES, J. 2007. Target mimicry provides a new mechanism for regulation of microRNA activity. *Nature Genetics,* 39**,** 1033-1037.

GALLART, A. P., PULIDO, A. H., DE LAGRÁN, I. A. M., SANSEVERINO, W. & CIGLIANO, R. A. 2016. GREENC: a Wiki-based database of plant lncRNAs. *Nucleic Acids Research,* 44**,** D1161-D1166.

GU, M., LIU, W., MENG, Q., ZHANG, W., CHEN, A., SUN, S. & XU, G. 2014. Identification of microRNAs in six solanaceous plants and their potential link with phosphate and mycorrhizal signaling. *Journal of Integrative Plant Biology,* 56**,** 1164-1178.

HEO, J. B. & SUNG, S. 2011. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science,* 331**,** 76-79.

JIN, J., LIU, J., WANG, H., WONG, L. & CHUA, N.-H. 2013. PLncDB: plant long noncoding RNA database. *Bioinformatics***,** btt107.

JONES, P., BINNS, D., CHANG, H.-Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A. & NUKA, G. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics,* 30**,** 1236-1240.

KIM, E.-D. & SUNG, S. 2012. Long noncoding RNA: unveiling hidden layer of gene regulatory networks. *Trends in Plant Science,* 17**,** 16-21.

KONG, L., ZHANG, Y., YE, Z.-Q., LIU, X.-Q., ZHAO, S.-Q., WEI, L. & GAO, G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Research,* 35**,** W345-W349.

KRZYWINSKI, M., SCHEIN, J., BIROL, I., CONNORS, J., GASCOYNE, R., HORSMAN, D., JONES, S. J. & MARRA, M. A. 2009. Circos: an information aesthetic for comparative genomics. *Genome Research,* 19**,** 1639-1645.

KUBHEKA, G. C., COUTINHO, T. A., MOLELEKI, N. & MOLELEKI, L. N. 2013. Colonization patterns of an mCherry-Tagged Pectobacterium carotovorum subsp. brasiliense Strain in potato plants. *Phytopathology,* 103**,** 1268-1279.

LECHNER, M., FINDEIß, S., STEINER, L., MARZ, M., STADLER, P. F. & PROHASKA, S. J. 2011. Proteinortho: Detection of (Co-) orthologs in large-scale analysis. *BMC Bioinformatics,* 12**,** 124.

LI, F., PIGNATTA, D., BENDIX, C., BRUNKARD, J. O., COHN, M. M., TUNG, J., SUN, H., KUMAR, P. & BAKER, B. 2012. MicroRNA regulation of plant innate immune receptors. *Proceedings of the National Academy of Sciences,* 109**,** 1790-1795.

LI, L., EICHTEN, S. R., SHIMIZU, R., PETSCH, K., YEH, C.-T., WU, W., CHETTOOR, A. M., GIVAN, S. A., COLE, R. A. & FOWLER, J. E. 2014. Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biology,* 15**,** R40.

LIU, J., JUNG, C., XU, J., WANG, H., DENG, S., BERNAD, L., ARENAS-HUERTERO, C. & CHUA, N.-H. 2012a. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *The Plant Cell Online,* 24**,** 4333-4345.

LIU, J., JUNG, C., XU, J., WANG, H., DENG, S., BERNAD, L., ARENAS-HUERTERO, C. & CHUA, N.-H. 2012b. Genome-wide analysis uncovers regulation of long intergenic noncoding RNAs in Arabidopsis. *The Plant Cell,* 24**,** 4333-4345.

LIU, X., HAO, L., LI, D., ZHU, L. & HU, S. 2015. Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics, Proteomics and Bioinformatics*.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods,* 25**,** 402-408.

LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology,* 15**,** 550.

MA, L., BAJIC, V. B. & ZHANG, Z. 2013. On the classification of long non-coding RNAs. *RNA Biology,* 10**,** 924-933.

MERCER, T. R., DINGER, M. E. & MATTICK, J. S. 2009. Long non-coding RNAs: insights into functions. *Nature Reviews Genetics,* 10**,** 155-159.

MI, H., POUDEL, S., MURUGANUJAN, A., CASAGRANDE, J. T. & THOMAS, P. D. 2016. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Research,* 44**,** D336-D342.

MITCHELL, A., CHANG, H.-Y., DAUGHERTY, L., FRASER, M., HUNTER, S., LOPEZ, R., MCANULLA, C., MCMENAMIN, C., NUKA, G. & PESSEAT, S. 2014. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Research***,** gku1243.

NICOT, N., HAUSMAN, J.-F., HOFFMANN, L. & EVERS, D. 2005. Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. *Journal of Experimental Botany,* 56**,** 2907-2914.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-842.

RHEE, S. Y. & MUTWIL, M. 2014. Towards revealing the functions of all genes in plants. *Trends in Plant Science,* 19**,** 212-221.

SAITO, K., HIRAI, M. Y. & YONEKURA-SAKAKIBARA, K. 2008. Decoding genes with coexpression networks and metabolomics–'majority report by precogs'. *Trends in Plant Science,* 13**,** 36-43.

SALDANHA, A. J. 2004. Java treeview-extensible visualization of microarray data. *Bioinformatics,* 20.

SHIN, H., SHIN, H. S., CHEN, R. & HARRISON, M. J. 2006. Loss of At4 function impacts phosphate distribution between the roots and the shoots during phosphate starvation. *The Plant Journal,* 45**,** 712-726.

SHIN, J. H. & CHEKANOVA, J. A. 2014. Arabidopsis RRP6L1 and RRP6L2 function in FLOWERING LOCUS C silencing via regulation of antisense RNA synthesis. *PLoS Genetics,* 10**,** e1004612.

SHUAI, P., LIANG, D., TANG, S., ZHANG, Z., YE, C.-Y., SU, Y., XIA, X. & YIN, W. 2014. Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in Populus trichocarpa. *Journal of Experimental Botany*, eru256.

SONG, D., YANG, Y., YU, B., ZHENG, B., DENG, Z., LU, B.-L., CHEN, X. & JIANG, T. 2009. Computational prediction of novel non-coding RNAs in Arabidopsis thaliana. *BMC Bioinformatics,* 10, S36.

SWIEZEWSKI, S., LIU, F., MAGUSIN, A. & DEAN, C. 2009. Cold-induced silencing by long antisense transcripts of an Arabidopsis Polycomb target. *Nature,* 462, 799-802.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25, 1105-1111.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols,* 7, 562-578.

ULITSKY, I. & BARTEL, D. P. 2013. lincRNAs: genomics, evolution, and mechanisms. *Cell,* 154, 26-46.

VAN DER MERWE, J. J., COUTINHO, T. A., KORSTEN, L. & VAN DER WAALS, J. E. 2010. Pectobacterium carotovorum subsp. brasiliensis causing blackleg on potatoes in South Africa. *European Journal of Plant Pathology,* 126, 175-185.

WANG, M., YUAN, D., TU, L., GAO, W., HE, Y., HU, H., WANG, P., LIU, N., LINDSEY, K. & ZHANG, X. 2015. Long noncoding RNAs and their proposed functions in fibre development of cotton (Gossypium spp.). *New Phytologist*.

WEN, J., PARKER, B. J. & WEILLER, G. F. 2007. In silico identification and characterization of mRNA-like noncoding transcripts in Medicago truncatula. *In silico Biology,* 7, 485-505.

WIERZBICKI, A. T. 2012. The role of long non-coding RNA in transcriptional gene silencing. *Current Opinion in Plant Biology,* 15, 517-522.

WOLFE, C., KOHANE, I. & BUTTE, A. 2005. Systematic survey reveals general applicability of "guilt-by-association" within gene coexpression networks. *BMC Bioinformatics,* 6, 227.

WU, H.-J., WANG, Z.-M., WANG, M. & WANG, X.-J. 2013. Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiology,* 161, 1875-1884.

XIE, F., FRAZIER, T. P. & ZHANG, B. 2011. Identification, characterization and expression analysis of MicroRNAs and their targets in the potato (Solanum tuberosum). *Gene,* 473, 8-22.

XIN, M., WANG, Y., YAO, Y., SONG, N., HU, Z., QIN, D., XIE, C., PENG, H., NI, Z. & SUN, Q. 2011. Identification and characterization of wheat long non-protein coding RNAs responsive to powdery mildew infection and heat stress by using microarray analysis and SBS sequencing. *BMC Plant Biology,* 11, 61.

YI, X., ZHANG, Z., LING, Y., XU, W. & SU, Z. 2015. PNRD: a plant non-coding RNA database. *Nucleic Acids Research,* 43, D982-D989.

ZHANG, R., MARSHALL, D., BRYAN, G. J. & HORNYIK, C. 2013. Identification and characterization of miRNA transcriptome in potato by high-throughput sequencing. *PLoS One,* 8, e57233.

ZHANG, W., LUO, Y., GONG, X., ZENG, W. & LI, S. 2009. Computational identification of 48 potato microRNAs and their targets. *Computational Biology and Chemistry,* 33, 84-93.

ZHANG, Y.-C. & CHEN, Y.-Q. 2013. Long noncoding RNAs: New regulators in plant development. *Biochemical and Biophysical Research Communications,* 436, 111-114.

ZHANG, Y.-C., LIAO, J.-Y., LI, Z.-Y., YU, Y., ZHANG, J.-P., LI, Q.-F., QU, L.-H., SHU, W.-S. & CHEN, Y.-Q. 2014. Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biology,* 15, 512.

ZHU, B., YANG, Y., LI, R., FU, D., WEN, L., LUO, Y. & ZHU, H. 2015. RNA sequencing and functional analysis implicate the regulatory role of long non-coding RNAs in tomato fruit ripening. *Journal of Experimental Botany***,** erv203.

ZHU, Q. H., STEPHEN, S., TAYLOR, J., HELLIWELL, C. A. & WANG, M. B. 2014. Long noncoding RNAs responsive to Fusarium oxysporum infection in Arabidopsis thaliana. *New Phytologist,* 201**,** 574-584.

Table 3.1: Comparison of lincRNA and protein-coding genes in *S. tuberosum* cvs Valor and BP1

	0	1	2	3	4	5	6	7	8	9	≥10	Total transcripts
Valor lincRNA	5	791	145	93	33	11	9	4	6	6	10	1113
BP1 lincRNA	17	779	158	74	33	11	14	5	9	3	10	1113
mRNA[a]	0	12415	7539	5363	3321	2476	1704	1355	1094	881	2880	39028
mRNA[b]	0	12105	7839	5495	3405	2467	1679	1356	1039	852	2791	39028

a exons from mRNA transcripts assembled in this study (merged from Valor and BP1)

b exons from mRNA transcripts based on potato genome annotation (PGCS_v4.03)

Table 3.2: LincRNA transcripts highly co-expressed with defense-related CDS genes

LincRNA ID	LincRNA class	r_{rho}*	Potato gene ID	Arabidopsis ortholog	Gene description	Gene Ontology classification (Biological process)	
LincRNA739	Intergenic	1	PGSC0003DMG400014801	AT4G17760	rad1-like	response to stimulus (GO:0050896)	response to stress (GO:0006950)
LincRNA1304	Intergenic	1	PGSC0003DMG400020345	AT2G38080	Laccase-4	response to stimulus (GO:0050896)	response to toxic substance (GO:0009636)
LincRNA127	Intergenic	1	PGSC0003DMG400011631	AT3G46230	17.4 kDa class I heat shock protein (HSP17.4A)	response to stimulus (GO:0050896)	response to stress (GO:0006950)
LincRNA803	Intergenic	1		AT1G53540	17.6 kDa class I heat shock protein 3 (HSP17.6C)	response to stimulus (GO:0050896)	response to stress (GO:0006950)
LincRNA127	Intergenic	1	PGSC0003DMG400000996	AT3G47570	Probable LRR receptor-like serine/threonine-protein kinase	response to stimulus (GO:0050896)	defense response to bacterium (GO:0042742)
				AT3G47580	Leucine-rich repeat protein kinase family protein		
				AT3G47090	Leucine-rich repeat protein kinase-like protein		
LincRNA1464	Intergenic	1	PGSC0003DMG400012994	AT5G21950	Hydrolase, alpha/beta fold family protein	response to stimulus (GO:0050896)	response to toxic substance (GO:0009636)
				AT4G33180			
LincRNA907	Intergenic	1	PGSC0003DMG400000757	AT2G23620	Methylesterase 1	response to stimulus (GO:0050896)	response to toxic substance (GO:0009636)
LincRNA1118	Intergenic	0.9					
LincRNA258	Intergenic	0.9					
LincRNA632	Intergenic	1	PGSC0003DMG400025635	AT3G45920	Protein kinase family protein		defense response to bacterium (GO:0042742)
			PGSC0003DMG400004885	AT2G24370	Adenine nucleotide alpha hydrolase domain-containing protein kinase		defense response to bacterium (GO:0042742)
LincRNA758	Intergenic	1					
LincRNA749	Intergenic	1					
				AT4G09570	Calcium-dependent protein kinase 4		response to endogenous stimulus (GO:0009719)
			PGSC0003DMG400026077	AT1G35670	Calcium-dependent protein kinase 11		
LincRNA758	Intergenic	1					
LincRNA749	Intergenic	1					
			PGSC0003DMG400013679	AT1G30270	CBL-interacting serine/threonine-protein kinase 23	response to stimulus (GO:0050896)	
LincRNA1112	Intergenic	1					
LincRNA908	Intergenic	0.9					
LincRNA1712	Intergenic	0.9					

			PGSC0003DMG400030755	AT1G77110	Probable auxin efflux carrier component 6	response to stimulus (GO:0050896)	response to endogenous stimulus (GO:0009719)
LincRNA178	Intergenic	0.9					
LincRNA1583	Intergenic	1					
			PGSC0003DMG400021008	AT4G20940	Probable LRR receptor-like serine/threonine-protein kinase	response to stimulus (GO:0050896)	defense response to bacterium (GO:0042742)
LincRNA379	Intergenic	0.9					

*Spearman correlation coefficient

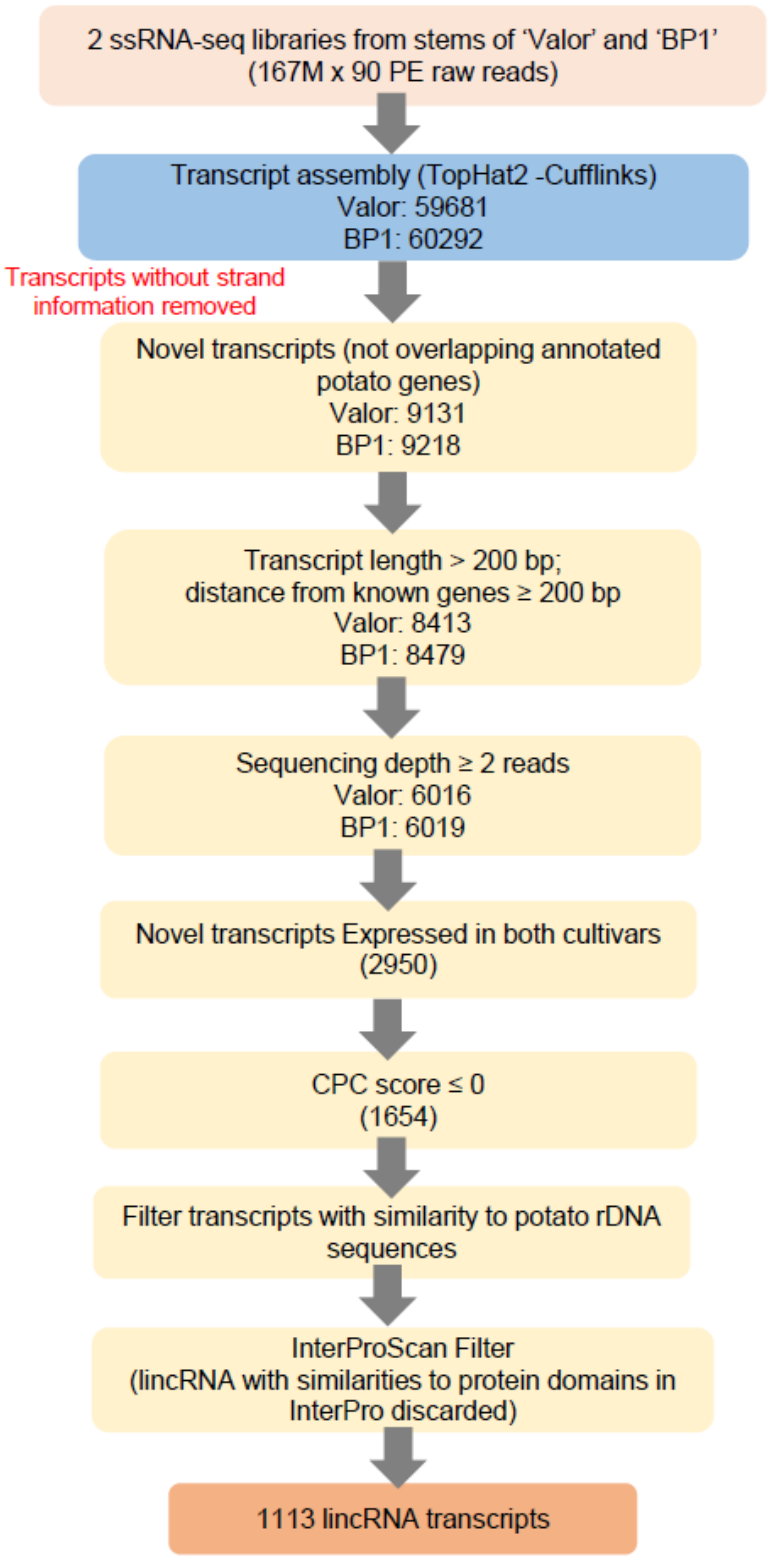


Fig. 3.1: Schematic diagram of the bioinformatics approach used for identification of potato lincRNAs

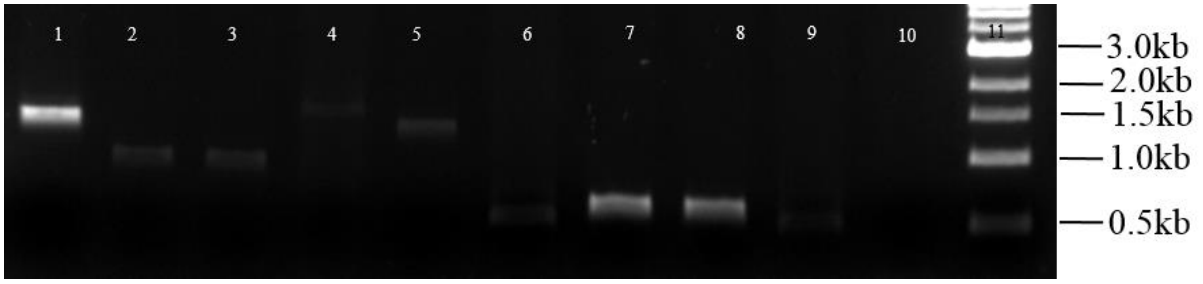


Fig. 3.2: RT-PCR validation of nine lincRNA transcripts. Agarose gel electrophoresis of the PCR amplicon fragments representing each lincRNA. Lane 1. LincRNA9, Lane 2. LincRNA10, Lane 3. LincRNA12, Lane 4. LincRNA13, Lane 5. LincRNA20, Lane 6. LincRNA178, Lane 7. LincRNA1405, Lane 8. LincRNA24, Lane 9. LincRNA1102, Lane 10. No reverse transcriptase control, Lane 11. 1 kb DNA ladder.

a

b

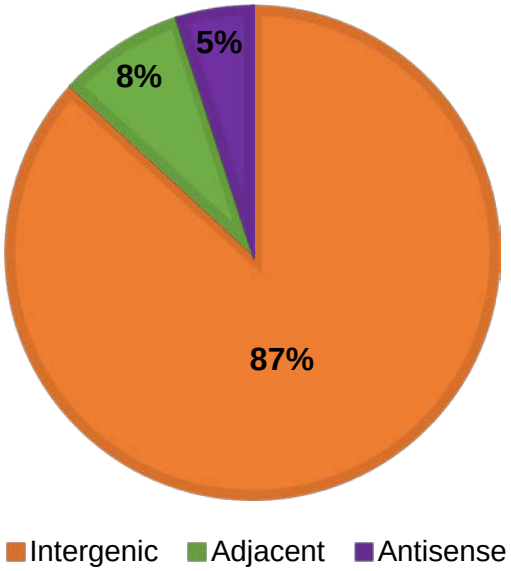
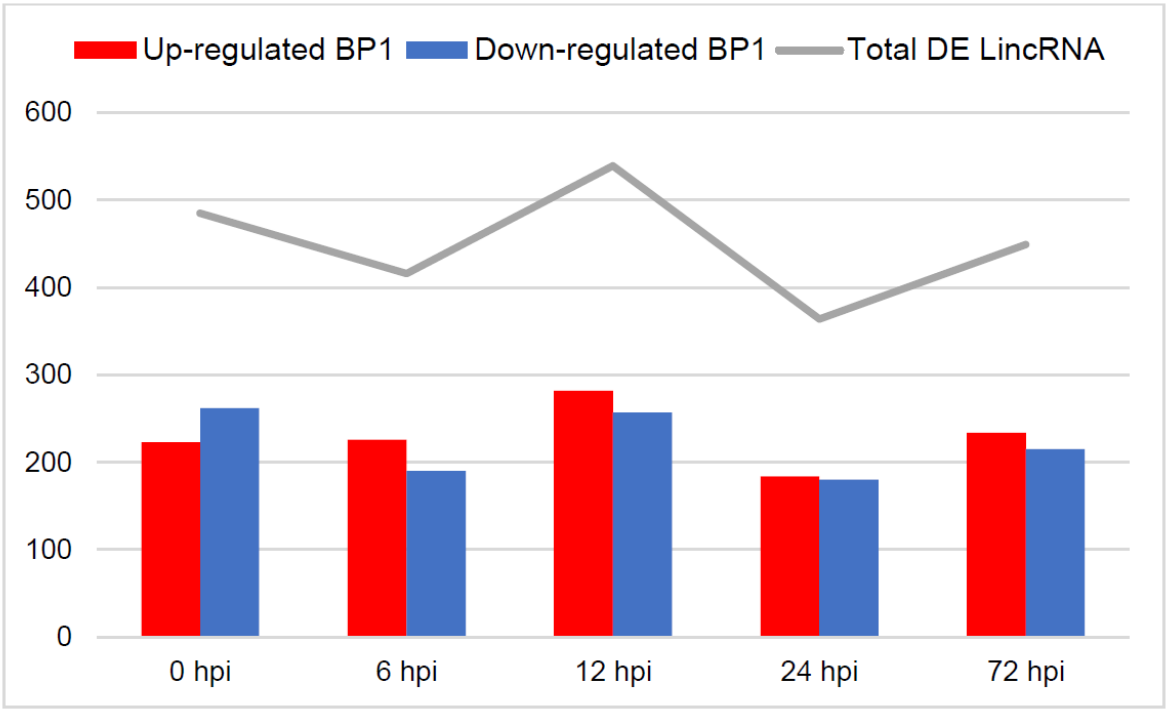
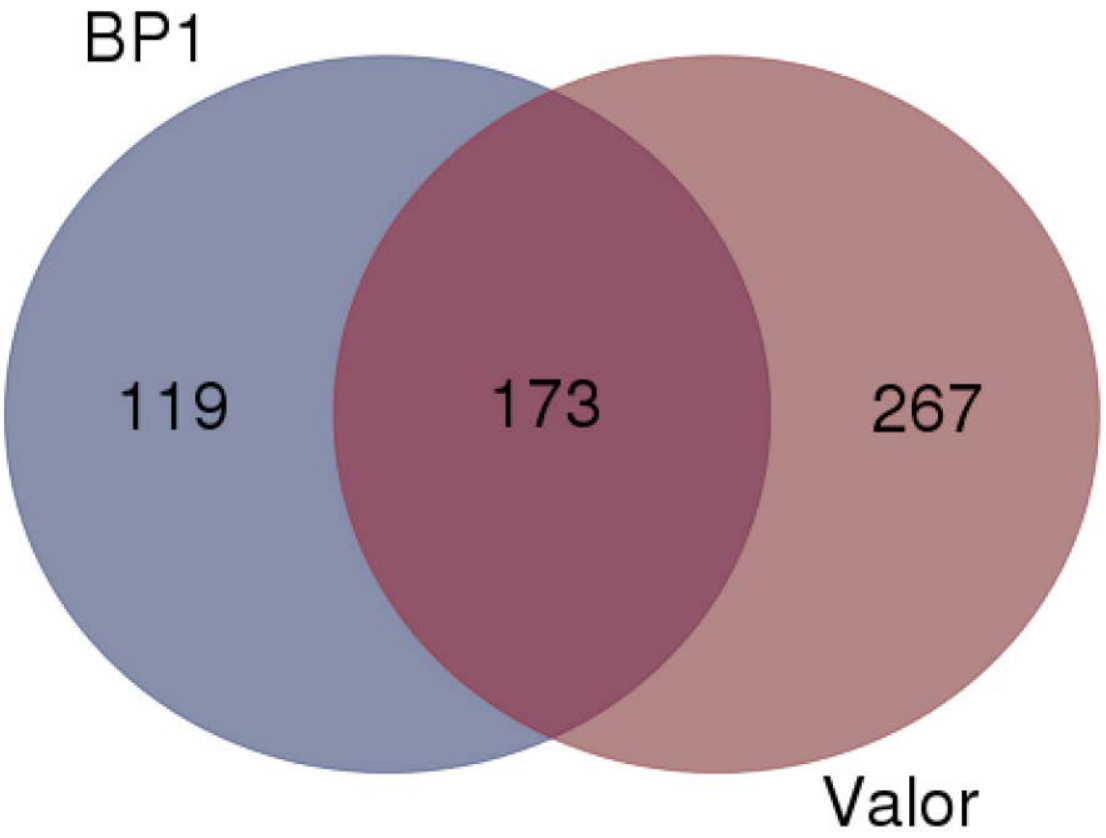


Fig. 3.3: Classification of potato lincRNAs relative to protein-coding transcripts. (**a**) A schematic diagram showing the location of lincRNAs in relation to adjacent protein-coding genes (black rectangles). Purple arrows represent antisense-lncRNAs which overlap annotated genes on the opposite strand; Green arrows show adjacent-lncRNAs which are positioned in close proximity to annotated genes; and the Orange arrow represents long intergenic noncoding RNAs (lincRNAs). (**b**) Percentage and distribution of lincRNAs in three classes.

a



b



c

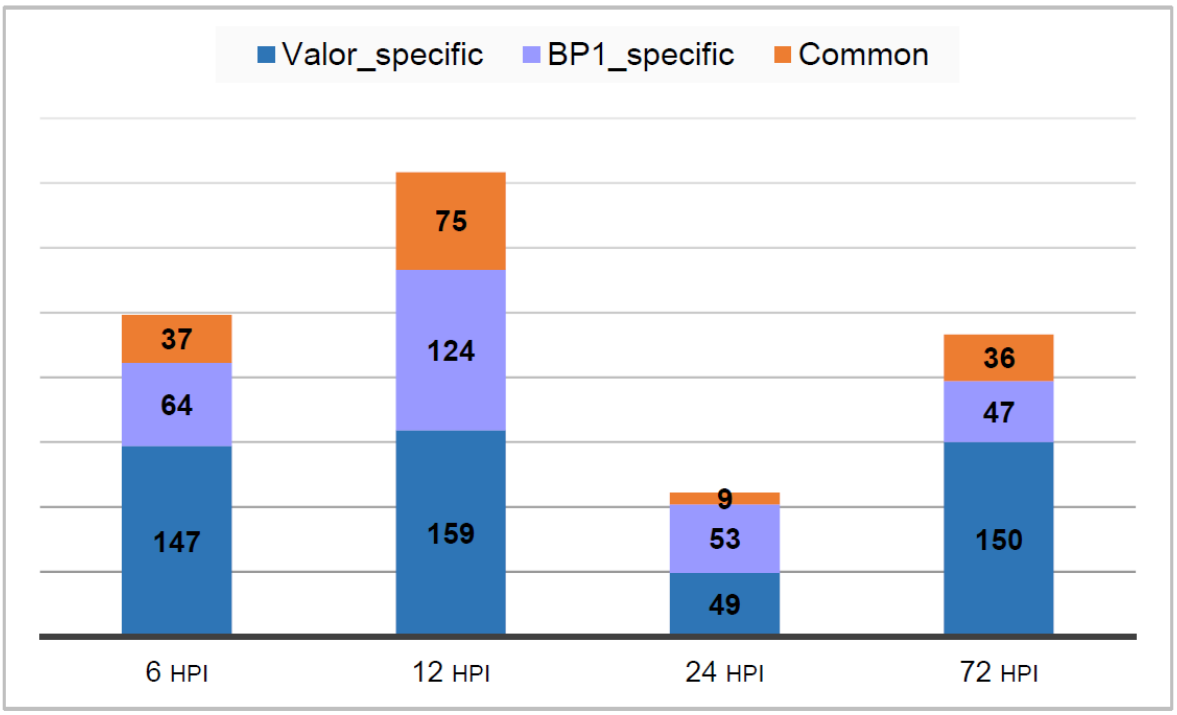


Fig. 3.4: LincRNAs significantly expressed over time between Valor and BP1 following infection with *Pcb1692*. (**a**) Pairwise comparisons between *S. tuberosum* cv Valor and *S. tuberosum* cv BP1 at each time-point.Red represent significantly upregulated and blue represent significantly downregulated. (**b**) Comparison of DE lincRNAs specific to each cultivar in relation to the mock-inoculated samples (0 hpi). (**c**) Numbers of DE lincRNAs common or specific to each cultivar at individual time-points in relation to mock inoculated samples.

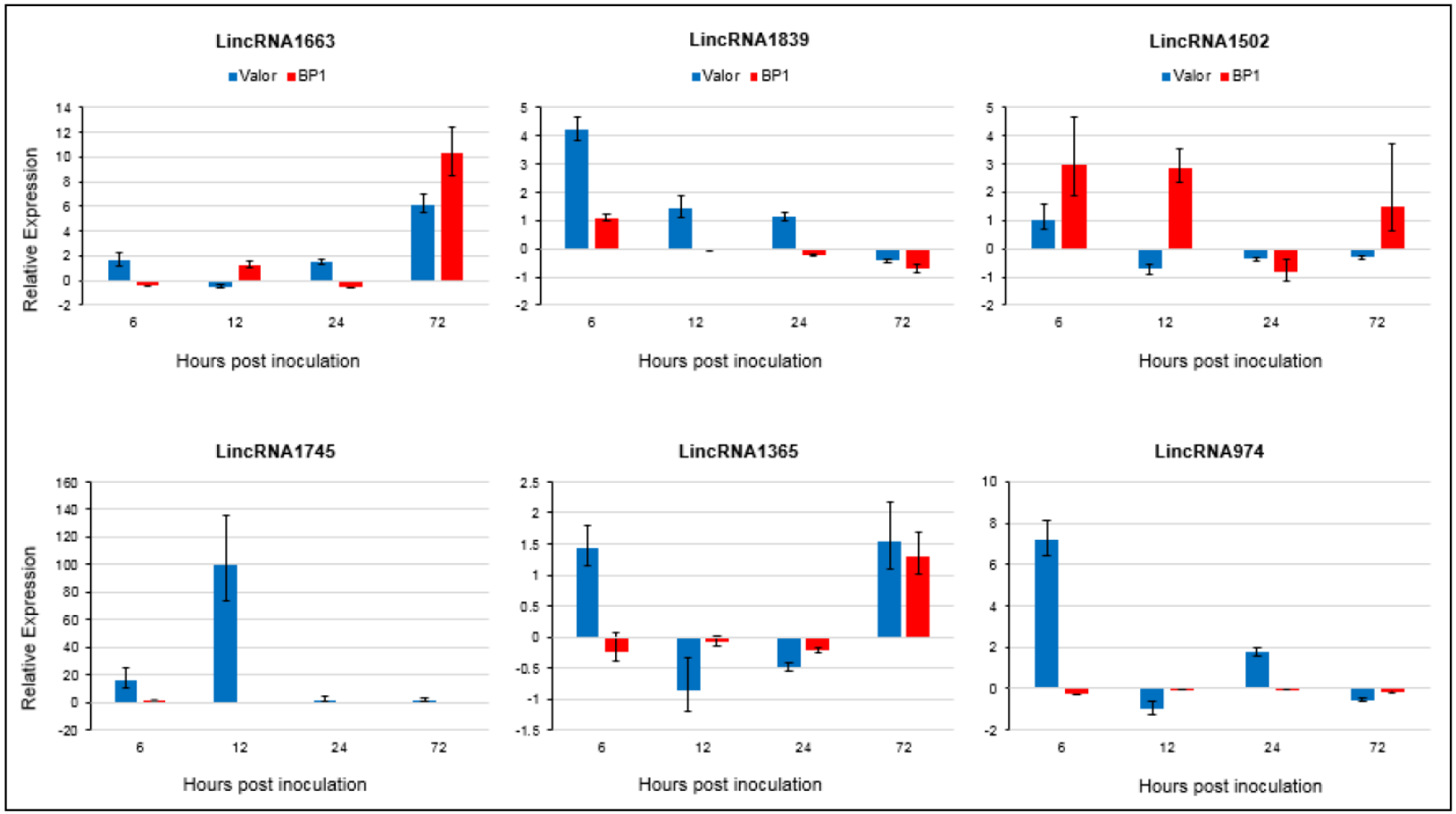
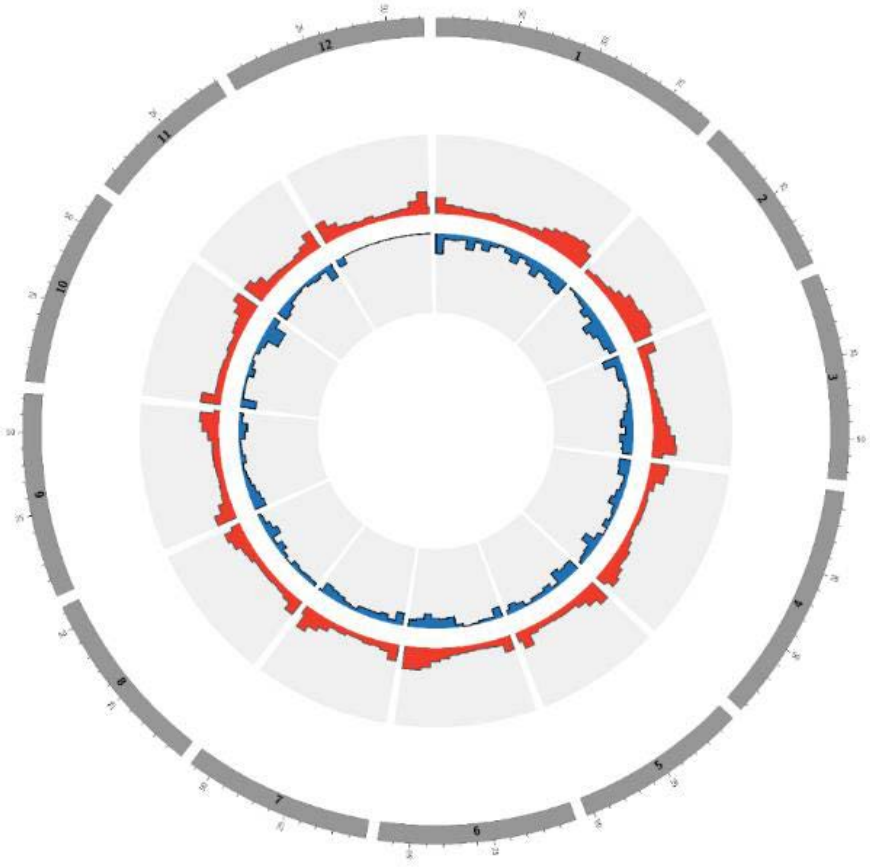


Fig. 3.5: RT-qPCR validation of time-course RNA-seq data using 6 selected lincRNAs differentially expressed over time. 18S rRNA and elongation factor 1-α (ef1α) were used as the reference genes. The relative expression levels of lincRNAs at each time point were calculated relative to calibrator (control sample; 0 hpi). Error bars represent the range of relative expression (fold change) calculated by $2^{-(\Delta\Delta Ct \pm SD)}$. Two biological replicates were used in triplicate.

a



b

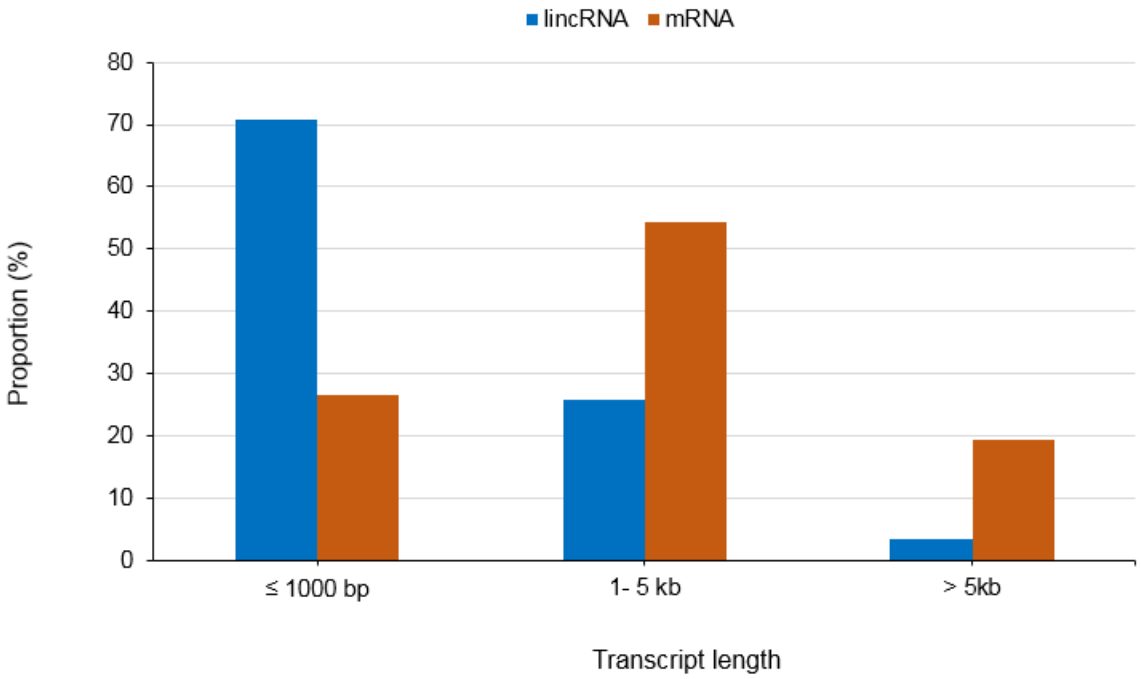


Fig. S3.1: (**a**) Comparison of the genomic distribution of lincRNAs and protein-coding genes across the 12 potato chromosomes. The outer grey track represents the 12 potato chromosomes, with a scale (Mb) showing the length of each chromosome. The red histograms (second track with an outer orientation) and blue histograms (third track with inner orientation) represent the abundance and distribution of mRNA and lincRNAs, respectively, throughout the potato genome. The bin size (histogram width) = 5 Mbp). (**b**) Comparison of LincRNA lengths to protein-coding mRNA transcripts in potato (PGSC_DM_v4.03 genome assembly).

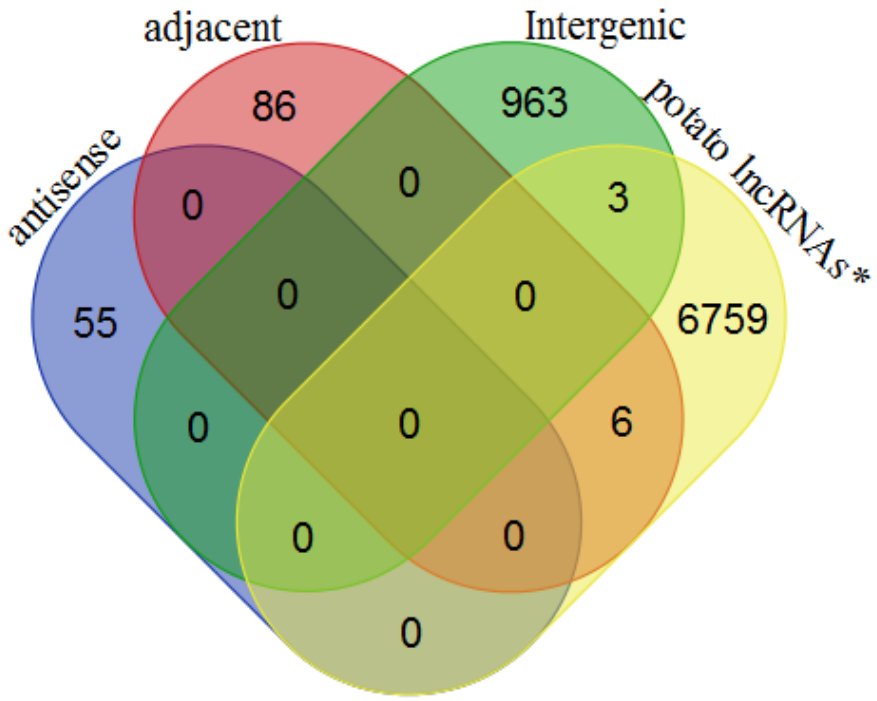


Fig. S3.2: Comparison of the 1113 lincRNA transcripts identified in the present study with potato lncRNAs available in the GreenC database.

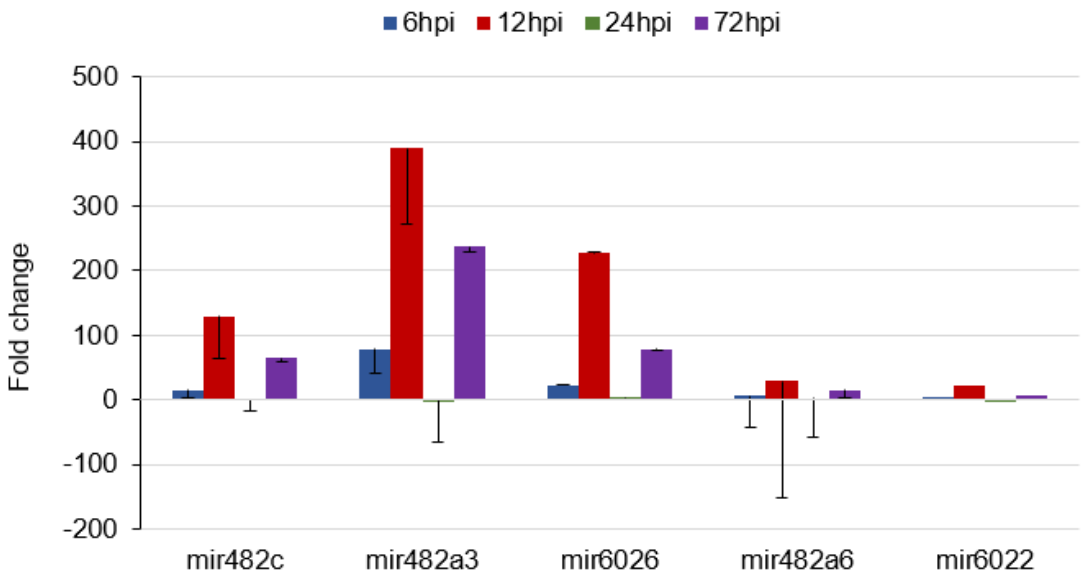


Fig. S3.3: RT-qPCR confirmation of five potato defense-related miRNAs in S. tuberosum cv BP1, computationally predicted to target some of the lincRNA transcripts. U6 snRNA was used as the reference gene. The fold changes of miRNAs at each time point were calculated relative to calibrator (control sample; 0 hpi). The experiments were done in triplicate. Error bars represent the fold change range calculated by 2-(ΔΔCt±SD).

CHAPTER FOUR

Discovery and profiling of small RNAs responsive to stress conditions in the plant pathogen Pectobacterium atrosepticum

Stanford Kwenda[1], Vladimir Gorshkov[2,3], Aadi Moolam Ramesh[1], Sanushka Naidoo[4], Enrico Rubagotti[5], Paul R. J. Birch[6], Lucy N. Moleleki[1]

[1]Forestry and Agricultural Biotechnology Institute (FABI), Genomics Research Institute (GRI), Department of Microbiology and Plant Pathology, University of Pretoria, Pretoria 0028, South Africa.

[2]Kazan Institute of Biochemistry and Biophysics, Kazan Scientific Center, Russian Academy of Sciences, Kazan, Russia.

[3]Department of Botany and Plant Physiology, Kazan Federal University, Kazan, Russia.

[4]Department of Genetics, Forestry and Agricultural Biotechnology (FABI), University of Pretoria, Pretoria, South Africa.

[5]Genomics Research Institute, Centre for Microbial Ecology and Genomics (CMEG), University of Pretoria, Pretoria, South Africa.

[6]The Division of Plant Sciences, College of Life Sciences, University of Dundee (The James Hutton Institute), Dundee DD25DA, Scotland, UK

This chapter has been prepared in the format of a manuscript and has been published in the peer-reviewed journal BMC Genomics (Kwenda *et al.*, BMC Genomics. (2016) Jan 12; 17:47. doi: 10.1186/s12864-016-2376-0). I conceived the study with Dr. Vladimir Gorshkov and Prof Lucy N. Moleleki, I carried out the bioinformatics data analyses, performed the experimental work, and wrote the manuscript. Dr. Aadi Moolam Ramesh and Dr. Sanushka Naidoo contributed to the experimental validation of novel sRNAs. Prof Paul R. J. Birch critically reviewed the manuscript. Prof Lucy, N. Moleleki, contributed in drafting the manuscript, and she obtained funding which supported this work.

4.1 Abstract

Background: Small RNAs (sRNAs) have emerged as important regulatory molecules and have been studied in several bacteria. However, to date, there have been no whole-transcriptome studies on sRNAs in any of the Soft Rot *Enterobacteriaceae* (SRE) group of pathogens. Although the main ecological niches for these pathogens are plants, a significant part of their life cycle is undertaken outside their host within adverse soil environment. However, the mechanisms of SRE adaptation to this harsh nutrient-deficient environment are poorly understood.

Results: In the study reported herein, by using strand-specific RNA-seq analysis and *in silico* sRNA predictions, we describe the sRNA pool of *Pectobacterium atrosepticum* and reveal numerous sRNA candidates, including those that are induced during starvation-activated stress responses. Consequently, strand-specific RNA-seq enabled detection of 137 sRNAs and sRNA candidates under starvation conditions; 25 of these sRNAs were predicted for this bacterium *in silico*. Functional annotations were computationally assigned to 68 sRNAs. The expression of sRNAs in *P. atrosepticum* was compared under growth-promoting and starvation conditions: 68 sRNAs were differentially expressed with 47 sRNAs up-regulated under nutrient-deficient conditions. Conservation analysis using BLAST showed that most of the identified sRNAs are conserved within the SRE. Subsequently, we identified 9 novel sRNAs within the *P. atrosepticum* genome.

Conclusions: Since many of the identified sRNAs are starvation-induced, the results of our study suggests that sRNAs play key roles in bacterial adaptive response. Finally, this work provides a basis for future experimental characterization and validation of sRNAs in plant pathogens.

4.2 Introduction

The importance of small RNAs (sRNAs) in bacterial gene expression regulation is now broadly appreciated (Waters and Storz, 2009, Wilms et al., 2012). sRNAs play essential regulatory roles in diverse processes including metabolic reactions, stress response, biofilm formation and pathogenesis (Peer and Margalit, 2011). They act as either activators or repressors of proteins and mRNAs. The length of most of the bacterial sRNAs ranges between 50 and 300 but can reach up to 500 nucleotides (Gottesman and Storz, 2011). The best studied bacterial regulatory sRNAs are those that act through base-pairing interactions with target RNAs, usually modulating gene expression post-transcriptionally by controlling the translation and stability of mRNAs. The majority of these are *trans*-acting sRNAs found within intergenic regions (IGRs). *Trans*-acting sRNAs typically regulate mRNAs encoded at different genomic locations on the chromosome in response to changes in environmental conditions (Waters and Storz, 2009). Furthermore, *trans*-encoded sRNAs tend to have limited complementarity with their target RNAs and require the RNA chaperone Hfq to facilitate their pairing with mRNA targets (Gottesman and Storz, 2011). In contrast, *cis*-encoded antisense RNAs (asRNAs), also referred to as naturally occurring RNAs, are expressed on reverse strands opposite to annotated genes and have extensive complementarity with their target mRNAs (Gottesman and Storz, 2011). Antisense RNAs are thought to play physiological roles such as repression of genes encoding potentially toxic proteins (Fozo et al., 2008). Additional roles of asRNAs include blocking the translation of mRNA transcripts encoded on the opposite strand and directing their RNAse III-mediated cleavage (Gottesman and Storz, 2011). Other important classes of sRNAs include 1) riboswitches (leader sequences), which form part of the mRNA they regulate and usually present in the 5` UTR regions; 2) sRNAs which interact with proteins and modify their activities by mimicking their RNA or DNA targets, and 3) sRNAs with intrinsic regulatory activities (Gottesman and Storz, 2011).

The advent of RNA-seq for the resolution of messenger and structural RNAs has facilitated the analysis of vast numbers of sRNAs with increased sensitivity (Croucher et al., 2009,

Graveley, 2008). An additional benefit of RNA-seq approaches is that information about the direction of transcription can be resolved using directional RNA-seq (strand-specific RNA-seq; ssRNA-seq). This information is important for the detection of non-coding (nc) RNAs as well as 5` and 3` untranslated regions (UTRs), antisense transcripts and determination of overlapping features within the genome (Croucher et al., 2009). Combining deep sequencing with computational (*in silico*) prediction methods is emerging as an important approach for sRNA detection in bacterial genome sequences (Soutourina et al., 2013, Cho et al., 2014).

Pectobacterium atrosepticum is an important plant pathogen belonging to the bacterial family *Enterobacteriaceae* (Bell et al., 2004). This pathogen causes major yield losses globally through blackleg disease on potato plants in the field and potato tuber soft rot diseases during post-harvest storage. Most of the information on pectobacteria concerns their interaction with plant hosts, and little is known about how these bacteria spend much of their life outside of the host (Charkowski et al., 2012). However, it is known that *P. atrosepticum* is able to utilize various adaptive programs that enable bacteria to survive under adverse conditions (Gorshkov et al., 2010, Gorshkov et al., 2009). In a previous study, we showed that realization of these programs under nutrient-deficient conditions (starvation) is coupled with an increased transcript abundance of stress responsive genes in *P. atrosepticum,* and bacterial cells undergo morphological and ultrastructural changes (Petrova et al., 2013). In the current study we have evaluated the possible participation of sRNAs in bacterial starvation-induced stress response.

Few experimental studies on sRNAs have been carried out in *P. atrosepticum*. A well-known regulatory sRNA in *P. atrosepticum* is *rsmB*. This sRNA binds the RsmA protein, which is a homologue of *Escherichia coli* CrsA, a carbon storage regulator, and modulates its activity. In *P. atrosepticum* the RsmA/*rsmB* system regulates the production of virulence factors (Cui et al., 2001, Liu et al., 1998, Mukherjee et al., 1996). Moreover, a regulatory RNA antisense to the *expI* gene transcript, which encodes the synthase of mediators of quorum sensing (acyl-homoserine lactones), was found recently in *P. atrosepticum* (Gogoleva et al., 2014).

In the present study, identification of sRNAs in the complete genome of *P. atrosepticum* SCRI1043 was undertaken using *in silico* prediction and experimental validation via strand-specific RNA-sequencing. Both true (and/or known) and potentially novel sRNA candidates expressed under starvation conditions were identified. Differential expression analysis indicated that many of these sRNAs increase in abundance during exposure of bacteria to starvation compared to rich medium conditions, suggesting an important role of sRNAs in the survival of *P. atrosepticum* cells during nutrient deficiency induced stress.

4.3 Results and Discussion

4.3.1 Strand-specific RNA-seq detection of *P. atrosepticum* sRNAs under starvation-conditions

For experimental detection of sRNAs in *P. atrosepticum* SCRI1043, we used a combination of *in silico* and directional whole-transcriptome cDNA sequencing (strand-specific RNA-seq) [Fig. 4.1]. The experimental approach for determination of sRNA in *P. atrosepticum* is outlined in Fig. 4.1A. A total of 27.4 and 26.1 million paired-end (PE) reads were obtained from nutrient rich and starvation conditions, respectively. By using SAMtools (Li et al., 2009), PE reads mapped to each strand were extracted. Thus, enabling visualization of the sequence (PE) read alignments on the genome in a strand-specific manner. Visual inspections enabled the identification of candidate sRNA transcripts by manually analysing the position of PE reads with respect to annotated coding DNA sequences (CDS). This can be a particularly powerful approach to identify sRNAs and resolve their genomic positions because reads that map to intergenic regions may represent previously unannotated transcriptionally active non-coding sRNAs (Perkins et al., 2009). Only sRNA candidates with a length between 50 to 500 nucleotides were considered to be true positive sRNAs candidates. This technique enabled identification of a total of 137 sRNA candidates expressed under starvation condition (Supplementary Table S4.1). These candidate sRNAs were classified into four distinct sRNA groups based on their position in relation to adjacent CDSs: IGR/ trans-encoded sRNAs, asRNA, 5` UTR (riboswitches), and 3` UTR sRNAs (Fig. 4.2). An *in silico* approach (described in the section below) was employed to determine the putative transcriptional start sites (TSS) of the identified 137 sRNAs and to resolve their 5` ends. Only predicted TSS with transcription factor binding sites were considered as bona fide promoters. Thus, using this filter, TSS were identified upstream of 118 sRNA genes (Supplementary Table S4.2).

4.3.2 Identification of 3` UTR encoded sRNAs

We identified 15 sRNAs encoded within the 3` UTR regions of mRNA (referred to in this study as 3` UTR sRNAs) (Fig. 4.2). It is now appreciated that sRNAs not only originate from intergenic regions as independent transcripts but are also transcribed from 3` regions of coding mRNA (Guo et al., 2014). These 3` UTR sRNAs are generated either by means of mRNA transcript processing or as primary transcripts from an internal promoter within the mRNA coding sequence as in the case of *dapZ* sRNA (Chao et al., 2012). Thus, based on how they are produced, 3` UTR encoded sRNAs can be divided into 2 groups, that are: 1) sRNAs transcribed from an independent promoter located inside the overlapping mRNA gene or 3` UTR region (Type 1); and 2) sRNAs which are originated from the processing of the parent mRNA (Type 2) (Miyakoshi et al., 2015). Hence we used our ssRNA-seq data to determine whether the identified 3` UTR embedded sRNAs are transcribed independently from the parent mRNA. Ten 3` UTR sRNAs were considered to be independently transcribed based on comparisons of sRNA and parent mRNA RPKM (reads per kilobase of transcript per million mapped reads) values and the presence or absence of an internal promoter (Table 4.1). To determine the putative 5` ends and fundamental types of the detected 3` UTR sRNAs based on their biogenesis, we extracted each sRNA sequence plus 200 nt upstream of the start position of each sRNA and performed promoter predictions using BPROM program (www.softberry.com). This approach led to the identification of 14 distinct putative promoter sites (transcriptional start sites; TSS) embedded within the coding or 3` UTR regions of the parent mRNA upstream of each 3` UTR sRNA gene (Table 4.1). In addition, transcription factor binding sites were also detected within the predicted promoter regions. Taken together, the presence of putative internal promoter sites upstream of sRNAs TSS and the predicted transcriptional factor binding sites for each promoter, strongly suggests that fourteen 3` UTR sRNAs are type 1. Nine of which were also differentially expressed compared to their parent mRNAs based on RPKM values, further indicating evidence of independent expression. The remaining sRNA reg_seq13 could be a product of mRNA processing, thus type 2 since no

internal promoters supported by transcription binding sites were predicted for this sRNA. Overall, since the sRNA 5` ends and subsequent TSSs were predicted computationally, we were not able to determine whether these sRNAs possessed the characteristic 5`-triphosphate (5`-PPP) cap common to type 1 sRNAs in this present study.

The 137 sRNAs identified using strand-specific RNA-seq approach were checked against known *P. atrosepticum* SCRI1043 non-coding RNA descriptions on the Rfam database (Nawrocki et al., 2014). For this analysis, all descriptions for tRNAs, rRNAs and CRISPR RNAs were excluded. This also served to assess the efficiency of the strand-specific RNA-seq method in detecting sRNA transcripts. In total 56.6% (47/83) of the known *P. atrosepticum* sRNAs in the Rfam database were identified using ssRNA-seq of cells cultured under starvation conditions (Fig. 4.1B and Supplementary Table S4.1).

4.3.3 Computational prediction of sRNA in the *Pectobacterium atrosepticum* genome

Even though ssRNA-seq is a powerful tool for identification of sRNAs, it might be subject to some limitations. For example, since the formation of particular sRNAs is highly dependent on culture conditions, it is not possible to unravel the whole pool of sRNAs that is encoded in the genome of the target microorganism within the frameworks of a given experiment. Consequently, a combination of experimental and computational identification of sRNA is often seen as a more comprehensive approach towards identification of sRNAs (Kulkarni and Kulkarni, 2007, Livny and Waldor, 2007). Hence, in addition to ssRNA-seq, an *in silico* sRNA analysis was performed according to computational methods implemented previously (Pichon et al., 2012), with some modifications (see Fig. 4.1C for a schematic representation of the computational prediction strategy).

An initial step towards *in silico* sRNA candidate disclosure consisted of identification of predicted rho-independent terminators (RITs) in the *P. atrosepticum* SCRI1043 genome. Since about 72% of known sRNAs located within IGRs possess a RIT, computational methods based on prediction of RIT signature sequences have emerged as valuable algorithms for the

detection of sRNA molecules (Pichon et al., 2012, Soutourina et al., 2013). In intergenic and antisense to annotated open reading frames (ORF)in the *P. atrosepticum* SCRI1043 genome we detected a total of 1598 putative terminators (including both canonical and non-canonical terminators' candidates) with the 'Greatest ΔG' i.e. the most negative ΔG (free Gibbs energy) value. From the 1598 putative sRNA identified, 1165 were filtered out and excluded from further analysis due to the fact that their RITs were located less than 60 nucleotides downstream from stop codons of preceding annotated ORFs within the same strand. This resulted in identification of 433 sRNA candidates of 226 - 248 nt in length (Supplementary Table S4.3). To be more confident about the accuracy of the rho-independent terminator based prediction strategy used, a second prediction tool (SIPHT) (Livny, 2012) was employed. Herewith, the filtered set of sRNA candidate signatures was compared against sRNA predictions for *P. atrosepticum* SCRI1043 from the SIPHT web interface by means of BLAST local pairwise alignments using the genomic similarity search tool YASS (Noé and Kucherov, 2005), with standard parameters. Each comparison was made on both regular and complementary strands separately. As a result, a total of 105 and 101 matches (E-value < 0.001) were identified, partially or fully overlapping, for the forward and complementary strands, respectively. This additional filtering step combining comparative genomics with RIT based predictions yielded 206 sRNA candidates in *P. atrosepticum* SCRI1043 (Supplementary Table S4.4). Similarly to sRNA detected using ssRNA-Seq, predicted sRNAs were further classified into five distinct sRNA groups based on their position in relation to adjacent CDSs (results not included).

4.3.4 Comparison of RNA-seq results with computational sRNA predictions

The 208 candidate sRNAs identified computationally were compared to the 137 sRNA transcripts identified using ssRNA-seq. Only 25 of the *in silico* predicted sRNA candidates were also identified by RNA sequencing (Table 4.2). Such an incomplete overlap between computational sRNA predictions and deep sequencing detection has been noted in previous studies (Cho et al., 2014, Soutourina et al., 2013, Vockenhuber et al., 2011, Wilms et al.,

2012). It is possible that the discrepancy observed here could be largely because experimental detection of sRNAs was restricted to sRNAs expressed under one condition, *viz* starvation. Hence, it may well be that increasing the number of conditions in which RNA is harvested could lead to bridging the gap between *in silico* predicted and ssRNAseq identified sRNAs. Lastly, the disparity could be due to the presence of false positive *in silico* predictions as well as the elimination of sRNAs associated with RITs in close proximity to CDS regions when using RIT identification based *in silico* predictions. Nonetheless, the lengths of the majority of the *in silico* predicted sRNA transcripts were comparable to the sizes deduced from the strand-specific RNA-seq sRNA detections for the confirmed sRNA candidates.

4.3.5 Functional annotation of RNA-seq detected sRNAs

To describe and assign biological functions to the 137 sRNAs detected by strand-specific RNA-seq (including those confirmed by *in silico* predictions), we used the Rfam database (version 11.0) (Burge et al., 2013) and the RNAspace platform (Cros et al., 2011). The RNAspace platform comprises a suite of ncRNA prediction tools. Similarity searches on the RNAspace platform were restricted to comparative analysis and homology searches using BLAST/ YASS (sequence homology tools) against the Rfam 10.0 seed database and three RNA motif search tools, DARN, ERPIN and INFERNAL. In total, 68 sRNAs representing true (and/or known), previously described sRNA sequences were assigned into 6 functional classes (E-value < 0.001), and these included: 1 ribozyme, 21 riboswitches (consisting of 6 types), 14 RNA elements (10 different types), 30 sRNAs (including 9 Hfq-binding sRNAs), 1 asRNA and 1 tmRNA (Table 4.3). Amongst these, we characterized 13 sRNA sequences which were previously uncharacterized within the *P. atrosepticum* genome by means of Blast (e-value<0.001) and secondary structure predictions using the RNAfold Webserver (Hofacker et al., 1994). No functional classes were assigned to the remaining 69 sRNAs computationally, suggesting that they could be potentially novel sRNA candidates in *P. atrosepticum*.

Most of the detected riboswitches in this study corresponded to thiamine pyrophosphate (TPP) (Thi-box) riboswitches (Table 4.3). Bacterial riboswitches are embedded within the leader sequences (5` UTR regions) of numerous metabolic genes and act by repressing or activating their cognate genes at the translational level in gram-negative bacteria (Nudler and Mironov, 2004). Most thiamin-regulated genes encode transporters in different bacterial organisms (Miranda-Ríos et al., 2001). For example, TPP riboswitches identified are present upstream of genes involved in potassium transport (*trkD*), amino acid biosynthesis (*argG*), and genes related to the biosynthesis of secondary metabolites (*menD*). Generally, TPP riboswitches are found upstream (5` UTR regions) of many genes key in metabolic processes which use TPP as a cofactor (Miranda-Ríos et al., 2001). In this study, we also detected other riboswitches other than TPP-type riboswitches, that include Flavin mononucleotide (FMN), glycine, lysine, yybP-ykoy and MOCO RNA motif riboswitches.

Some of the detected RNA elements (leader sequences) were located upstream of operons or genes involved in biosynthesis of amino acids including leucine, histidine and tryptophan biosynthesis; and polysaccharide synthesis (Supplementary Table S4.1). It therefore seems plausible that most of the detected *cis*-regulatory elements are engaged in regulating processes involving substrate transport and biosynthesis in *P. atrosepticum*.

4.3.6 Conservation analysis of predicted sRNAs

The vast majority of known sRNAs are typically highly conserved across genera (Chen et al., 2011). We therefore analysed the conservation of identified sRNAs in *P. atrosepticum* SCRI1043 in five soft rot *Enterobacteriaceae* species whose complete genome sequences are available on GenBank. The 68 true/ known sRNA sequences with assigned functional classes were used for the conservation analysis. BLASTn analysis (E-value < 0.0001) using the YASS tool against *P. carotovorum* subsp. *carotovorum* PC1, *P. wasabiae* WPP163, *Pectobacterium spp.* SCC3193, *Dickeya dadantii* Ech703 and *D. zeae* Ech1591 complete genome sequences revealed that most sRNAs are conserved within the soft rot bacterial

species with 42 sRNAs (including 13 *trans*-encoded sRNAs, 18 riboswitches, 10 RNA elements, and 1 asRNAs) being present in all five SRE species (Fig. 4.3 and Supplementary Table S4.5). The high conservation of sRNAs within the SRE species emphasizes their regulatory importance in these bacteria. Six IGR sRNAs were conserved only in the *Pectobacterium* genus and belonged to two RNA families; namely *styR-44*, and *crcB* RNA motif (fluoride riboswitch) sensing fluoride ions and regulating the *crcB* gene (hypothetical protein) which possibly encodes a protein that functions by removing excess fluoride ions from the cell.

To be more confident with the 69 potentially novel sRNA candidates detected by ssRNA-seq, we filtered and screened them by checking their conservation within the five representative SRE strains using sequence similarity analysis. Nine of these candidate sRNAs had high sequence conservation (100% identity and coverage) within SRE strains and only single hits from the BLAST analysis and therefore were considered as novel sRNAs (Table 4.4). To validate the expression and lengths of the nine novel sRNAs, reverse transcription PCR (RT-PCR) was performed on cDNA of bacteria cells cultured under starvation conditions (Fig. 4.4). For each of the cDNA samples, a single amplicon that corresponded to the sRNA transcript size identified by ssRNA-seq was observed. As an additional validation step, the nucleotide bases of observed amplicons were confirmed by sequencing and alignment to respective sRNA sequences (Supplementary Table S4.6).

4.3.7 Differential expression of sRNAs under nutrient-rich and starvation conditions.

Application of strand-specific RNA-seq to study the transcriptome of *P. atrosepticum* uncovered an abundance of sRNAs including antisense transcripts, intergenic sRNAs and *cis*-encoded regulatory elements. The number of RNA-seq reads mapping to individual sRNA sequences provides a realistic assessment of relative transcript abundance (Arnvig et al., 2011), thus enabling quantification of differential expression of the sRNA transcripts in *P. atrosepticum* cells existing under nutrient-rich and nutrient-deficient (starvation) conditions.

The differential expression of sRNAs when growth conditions are changed could suggest potential functions and clarify conditions that induce or repress formation of specific sRNAs (Arnvig and Young, 2012). Hence, in order to understand the expression profiles of sRNAs in response to carbon and phosphorus starvation, we compared expression patterns of *P. atrosepticum* cells under nutrient-rich and nutrient-deficient (starvation) conditions. Based on the combined statistics of edgeR package (Robinson et al., 2010) (dispersion=0.04; q-value<0.1), and Gfold algorithm (v.1.1.4) (Feng et al., 2012), which uses a posterior distribution of log fold change for determining expression changes in experiments with single biological replication, thus, overcoming the shortcomings of relying on statistics based on p-value when biological replication is lacking (Feng et al., 2012). Subsequently, only sRNAs with significant differential expression from edgeR and Gfold analyses were considered. Thus, a total of 68 sRNA candidates were differentially expressed (Supplementary Table S4.7). Of these, 47 sRNAs were up-regulated under nutrient-deficient conditions (Table S4.7) suggesting that they are likely involved in regulatory mechanisms of stress response or adaptation in *P. atrosepticum*. To validate expression profiles identified by ssRNA-seq, we performed reverse transcription quantitative PCR (RT-qPCR) using three biological replicates, on eight selected sRNAs that were differentially expressed under nutrient-rich and starvation conditions. The RT-qPCR results confirmed expression patterns of these eight sRNA transcripts and validated our RNA-seq data (Fig. 4.5). Selected examples are discussed below.

We noticed that *rprA* was up-regulated (~1.5-fold) in *P. atrosepticum* under carbon-starvation conditions (Fig. 4.5 and Supplementary Table S4.7). The sRNA *rprA* acts by increasing (positively regulating) the translation of *rpoS* gene transcript (Repoila et al., 2003, Majdalani et al., 2001). RpoS is a sigma factor that controls the expression of stress responsive genes in bacteria during adverse conditions and stationary phase. We observed that the expression of *rpoS* gene in *P. atrosepticum* was higher during starvation than under nutrient-rich conditions (data not shown). This observation is consistent with previous data demonstrating

that RpoS is a principal regulator of the general stress response in bacteria allowing cells to survive environmental challenges as well as prepare for subsequent stresses (Hengge-Aronis, 2002). This is also consistent with our previous observations, demonstrating that *rpoS* gene expression increases significantly in *P. atrosepticum* under stress conditions (Petrova et al., 2013). Generally, the regulation of *rpoS* gene expression is known to be modulated at the translational level by at least four sRNA, namely, *arcZ*, *dsrA*, *rprA*, and *oxyS* in response to temperature, osmotic shock, oxidative stress and nutrient deprivation in *E. coli* (Repoila et al., 2003, Mandin and Gottesman, 2010). Hence, increased *rprA* expression observed in our study in *P. atrosepticum* under nutrient starvation conditions is likely to promote the enhanced translation of *rpoS* mRNA during adaptation of bacteria to starvation conditions.

ryhB2, a 106 nucleotide paralogue of *ryhB* sRNA, was up-regulated by a 15-fold magnitude in *P. atrosepticum* under nutrient-starvation conditions (Fig. 4.5). Generally, *ryhB,* regulates iron metabolism, including its acquisition and assimilation. *ryhB* acts by down-regulating expression of genes encoding iron-storage and iron-using proteins when iron is in limited supply. The main target genes for *ryhB* include the *sdhCDAB* operon encoding succinate dehydrogenase and *sodB* which encodes the iron-dependent superoxide dismutase (Richards and Vanderpool, 2011). *ryhB* expression level is usually inversely correlated with expression levels of the mRNA for the *sdhCDAB* operon (Hoe et al., 2013). This is consistent with our observations for *P. atrosepticum*: the transcription of the *sdhCDAB* operon was reduced under starvation conditions compared to the growth-promoting ones (Fig. 4.6).

Starvation conditions also induced the expression of *glmZ* (~2-fold increase) and *glmY* (5-fold increase) sRNAs in *P. atrosepticum* (Fig. 4.5). In enteric bacteria, these two sRNAs regulate amino sugar metabolism by activating the expression of *glmUS* operon which encodes the glucosamine-6-phosphate synthase, an essential enzyme in amino sugar metabolism (Urban and Vogel, 2008). The regulation by these two sRNAs modulates the transitions between carbon storage and carbon metabolism (Bobrovskyy and Vanderpool, 2013). The level of *glmY* is increased in the absence of glucosamine-6-phosphate leading to stabilization of *glmZ*. The

latter, in turn, activates *glmS* gene expression in an anti-antisense mechanism (Bobrovskyy and Vanderpool, 2013). GlmS enables cells to utilize the intermediates of glycolytic pathway including the fructose-6-phosphate for production of amino sugars. The glucosamine-6-phosphate is an essential precursor for biosynthesis of essential components of the cell envelope such as peptidoglycan and lipopolysaccharide in gram-negative bacteria. Thus, induction of *glmY* and *glmZ* expression in *P. atrosepticum* under starvation conditions likely indicates the important role of the amino sugar metabolism in adaptive response on this bacterium.

In summary, we have shown that several sRNAs are induced under nutrient-deficient compared to nutrient-rich conditions. We have also shown that induction of these sRNA leads to induction of various genes that potentially play a role in the survival of *P. atrosepticum*. In other members of the *Enterobacteriaceae* family including *E. coli* and *Salmonella*, sRNAs have also been shown to play an important role in adaptation to nutrient limited condition (Mika and Hengge, 2013). In these bacteria, sRNAs provide a signal that triggers production of extracellular polysaccharides (EPS) which in turn are involved in biofilm formation (Mika and Hengge, 2014). Although *P. atrosepticum* does not readily form biofilms *in vitro*, the overexpression of a diguanylate cyclase (PleD*), induced formation of biofilms suggesting that biofilm formation in this pathogen is cryptic and can be activated under optimum conditions (Pérez-Mendoza et al., 2011). Part of the pathogenesis of *P. atrosepticum* is in xylem tissue (when causing black leg disease of potato stems). The xylem is typified by limited nutrients and as such a harsh environment that requires well defined methods of survival. Hence, it is not surprising that many xylem dwelling phytopathogens such as *Xanthomonas*, *Clavibacter*, *Ralstonia* and *Xylella* form biofilms in xylem tissues of their respective hosts. Thus, it is possible that sRNA are extensively involved in the adaptation of *P. atrosepticum* and survival in stem vasculature. Identification of this suite of sRNA will allow us to study the role that these play in survival of this phytopathogen during stem colonisation.

4.4 Conclusions

In conclusion, in this study we have used a combination of strand-specific RNA-sequencing and *in silico* approaches to detect and analyse sRNAs in *P. atrosepticum* SCRI1043. We demonstrated the efficiency of ssRNA-seq in detecting sRNAs and determining the sRNA expression levels in response to specific bacterial growth conditions. A total of 137 sRNAs and sRNA candidates were experimentally detected in this study. We successfully determined sRNAs (that are riboswitches, *trans*-encoded sRNAs, 3` UTR sRNAs and asRNAs) that may play key roles in regulating stress responses. Most of the identified sRNAs in *P. atrosepticum* are conserved within the soft rot enterobacteria (SRE) species suggesting their importance in physiological responses for the SRE species. To our knowledge, this study constitutes the first genome/ transcriptome-wide analysis aimed at the discovery of sRNAs responsive to nutrient-deficiency (starvation) in bacteria. A significant fraction of the unravelled sRNAs appeared to be starvation responsive indicative of their importance in adaptation of bacteria to stress conditions. Determining the biological roles of these sRNAs will broaden our understanding of the diverse regulatory mechanisms they provide in modulating gene expression in *P. atrosepticum* and other SRE species during adaptation to changing environments.

4.5 Materials and Methods

4.5.1 Bacterial strains, media and culture conditions

A strain of *P. atrosepticum* SCRI1043 (Bell et al., 2004), was used in this study. sRNA profile was analysed in bacterial cells existing under growth-promoting and starvation conditions. The cultures with inoculation titer of 2–3 x 10^6 CFU (colony forming units) per ml were grown in Luria-Bertani medium (Sambrook et al., 1989), with aeration (200 r.p.m.) at 28 °C for 16 hours (growth-promoting conditions). Aliquots of these cultures were used for total RNA extraction. The remaining cells were transferred (after double wash) to carbon and phosphorus deficient AB medium containing 1 g l^{-1} NH_4Cl; 0.62 g l^{-1} MgSO4 x $7H_2O$; 0.15 g l^{-1} KCl; 0.013 g l^{-1} $CaCl_2$ x $2H_2O$ and 0.005 g l^{-1} $FeSO_4$ x $7H_2O$, pH 7.5 and incubated under starvation conditions with initial cell density of 5.4 x 10^8 ± 6.1 x 10^7 CFU per ml in glass vials without aeration at 28 °C (Petrova et al., 2014). Total RNA was extracted from 24 hour starving cells.

4.5.2 Total RNA preparation

Total RNA was isolated from bacterial cells using the RNeasy Protect Bacteria Mini Kit (Qiagen, USA), according to the manufacturer's instructions. Contaminating DNA was removed from the samples by DNAse (Qiagen) treatment. RNA was quantified using a Qubit fluorometer (Invitrogen, USA).

4.5.3 cDNA library construction and bacteria strand-specific RNA sequencing

Library construction and strand-specific sequencing were carried out at the Beijing Genomics Institute (BGI-Shenzhen, China; http://www.genomics.cn/en/index), following the manufacturer's protocols. Briefly, the rRNA was depleted from 1 microgram of total RNA using the Ribo-Zero Magnetic Gold Kit (Epicenter). TruSeq RNA Sample Prep Kit v2 (Illumina) was used for library construction. RNA was fragmented into small pieces using Elute Prime Fragment Mix. First-strand cDNA was synthesized with First Strand Master Mix and Super Script II (Invitrogen) reverse transcription (25°C for 10 min; 42°C for 50 min; 70°C for 15 min).

After product purification (Agencourt RNAClean XP Beads, AGENCOURT) the second-strand cDNA library was synthesized using Second Strand Master Mix and dATP, dGTP, dCTP, dUTP mix (1 hour at 16°C). Purified fragmented cDNA was end repaired (30 min at 30°C) and purified with AMPureXP Beads (AGENCOURT). Addition of the poly (A) tail was done with A-tailing Mix (30 min at 37°C) prior to ligating sequencing adapters (10 min at 30°C). The second-strand cDNA was degraded using the Uracil-N-Glycosylase (UNG) enzyme (10 min at 37°C) and the product purified by AMPureXP Beads (AGENCOURT). Several rounds of PCR amplification with PCR Primer Cocktail were performed to enrich the cDNA fragments and the PCR products were purified with AMPureXP Beads (AGENCOURT). Sequencing was performed using the Illumina HiSeq™ 2000 platform with pair-end 90 base reads.

4.5.4 Sequence read processing and experimental detection of sRNAs

Prior to analyzing the sequencing reads, adaptors were removed and the Illumina pair-end reads were quality checked using FASTQC:Read QC and trimmed using Trim sequences (version 1.0.0) implemented within the Galaxy software (Blankenberg et al., 2010, Giardine et al., 2005, Goecks et al., 2010). Quality trimmed reads were mapped to the *P. atrosepticum* SCRI1043 genome (http://www.ncbi.nlm.nih.gov/nuccore/50118965?report=fasta) using Bowtie2 (Langmead and Salzberg, 2012). The mapped reads in SAM format were converted to sorted and indexed BAM files using SAMtools version 0.1.18 (Li et al., 2009). Each BAM file was split into two separate forward and reverse strand alignments using SAMtools to obtain transcriptional direction. For visualization of the data in a strand-specific manner, the genome browser Artemis (Rutherford et al., 2000), was used. The strand-specific RNA Sequencing data from this study have been deposited in NCBI's Gene Expression Omnibus (GEO) with the accession number GSE68547.

4.5.5 RT-PCR validation of novel sRNA candidates

For RT-PCR, first-strand cDNA was synthesized from 1 µg of total RNA using Superscript™ III First-Strand cDNA Synthesis SuperMix kit according to the manufacturer's protocol

(Invitrogen, USA). The first-strand cDNA samples were used for RT-PCR, which was performed on Bio-RAD T100^{TM} Thermal Cycler conventional PCR (Bio-RAD, USA). To check for genomic DNA contamination, a non reverse-transcriptase control was included. The sRNA primers were designed online using Primer3 (Supplementary Table S4.8). PCR was performed in a 25 µl reaction mix containing 1 µl of template cDNA (~40 ng), Taq DNA Polymerase, 10x Taq Buffer (New England Biolabs, UK), 2.5 mM dNTPs each and 0. 5µM of forward and reverse primer each. Thermal cycling conditions were: 95 °C for 2 min; 30 cycles of 95 °C for 30 sec, 57 °C for 30 sec, 72 °C for 60 sec, and the final extension at 72°C for 5 min. The PCR products were analysed on 1.5% agarose gel including 100 bp DNA molecular weight ladder (NEB, UK).

4.5.6 Differential expression analysis of sRNAs

Artemis genome browser was used to create features of the discovered sRNAs on the *P. atrosepticum* reference genome and to make read counts for reads aligning to each strand under each growth condition. The read counts were used as input for the sRNA differential expression analysis using edgeR (Robinson et al., 2010). sRNA transcripts were considered differentially expressed provided that the p-value was < 0.05 and q-value < 0.1.

4.5.7 RT-qPCR validation of RNA-seq data

First strand cDNA synthesis was performed individually from total RNA samples from each of three biological replicates per condition using Superscript III First-Strand cDNA Synthesis SuperMix kit (Invitrogen, USA). For RT-qPCR, 2 µl of sample was added to 8 µl of Applied Biosystems SYBR Green Master Mix including each primer at a concentration of 0.4 µM and the reaction performed in the QuantStudio 12K Flex Real-Time PCR system (Life Technologies, USA). The following cycling conditions were used: an initial denaturation at 50 °C for 5 min and 95 ^{o}C for 2 min, followed by 45 cycles of 95 °C for 15 s and 60 °C for 1 min. Each sample was run in triplicate. Relative expression was measured using the comparative 2-$\Delta\Delta^{ct}$ method (Livak and Schmittgen, 2001) after normalizing the samples to

recA as the reference gene. Primers were designed using Primer3Plus (http://primer3plus.com/cgi-bin/dev/primer3plus.cgi) (Supplementary Table S4.9).

4.5.8 Soft rot bacteria genome sequences

The genome sequences of six soft rot bacteria species (*Pectobacterium atrosepticum* SCRI1043, *P. carotovorum* subsp. *carotovorum* PC1, *P. wasabiae* WPP163, *Pectobacterium* sp. SCC3193, *D. dadantii* Ech703, and *Dickeya zeae* Ech1591 were obtained from the European Nucleotide Archive (http://www.ebi.ac.uk/ena/).

4.5.9 Identification of RITs

The WebGeSTer DB (Mitra et al., 2011), database was used in this study to predict Rho-independent terminators (RITS) in *P. atrosepticum* SCRI1043 using default parameters. Briefly, no more than three mismatches were permitted within the stem structure and only RIT candidates with the highest ΔG score (ΔG <= -12.0 kcal/mol) were considered. Coordinates for putative RITs were obtained from the WebGeSTer DB, and java scripts were used to extract the sequences 200 nt upstream of the terminators (including the stem loop and tail sequences of the terminator). These sequences were considered as putative sRNA candidates and used in downstream sRNA prediction analysis. Additionally, known sRNAs within the *P. atrosepticum* SCRI1043 genome sequence were searched in the SIPHT web interface (published annotations) (Livny, 2012).

4.5.10 sRNA conservation analysis

The conservation of sRNA sequences detected using deep sequencing was determined by similarity analysis against sequences of complete genomes of five soft rot Enterobacteriaceae species using YASS (Noé and Kucherov, 2005), a sequence similarity search tool, with standard parameters.

4.5.11 Classification of sRNA

Following the model implemented for *Escherichia coli* (Pichon et al., 2012), custom scripts written in java were used to classify the predicted sRNA candidates into five non-coding RNA groups based on their position in relation to adjacent CDSs. Briefly, the first nucleotide in each RIT was used as the representative position of each sRNA candidate. To determine asRNA, the reference nucleotide on the opposite DNA strand had to be at least +15 nt relative to the ATG codon to -50 nt with respect to the stop codon. For 5` UTR, sRNA candidates had to be on the same DNA strand as the CDS and in a distance of < -100 nt upstream the ATG codon and for 3` UTR between +60 and +200 nt downstream of the stop codon. The rest of the remaining putative sRNAs were considered as IGR candidates if they were outside a CDS.

4.6 Supplementary Data

Table S4.1: Complete list of RNA-seq detected sRNAs

Table S4.2: Predicted transcription start sites of RNA-seq detected sRNAs

Table S4.3: List of *in silico* predicted sRNAs using RITs from WebGester DB. S2A: Forward strand predictions. S2B: Complementary strand predictions

Table S4.4: Combined list of predicted sRNA using SIPHT and RITs from WebGester DB. S3A: Matches of *in silico* predictions with SIPHT (forward strand) S3B: Matches of in silico predictions with SIPHT (complementary strand)

Table S4.5: Conservation analysis in Soft Rot *Enterobacteriaceae*

Table S4.6: Confirmation of RT-PCR amplicons by sequencing and BLASTn against respective sRNA sequences

Table S4.7: Differentially expressed sRNA under nutrient-rich and starvation conditions

Table S4.8: List of primers used for RT-PCR validation of novel sRNAs

Table S4.9: List of primers used for RT-qPCR validation of RNA-seq expression data

4.7 References

ARNVIG, K. & YOUNG, D. 2012. Non-coding RNA and its potential role in Mycobacterium tuberculosis pathogenesis. *RNA Biology,* 9**,** 427-436.

ARNVIG, K. B., COMAS, I., THOMSON, N. R., HOUGHTON, J., BOSHOFF, H. I., CROUCHER, N. J., ROSE, G., PERKINS, T. T., PARKHILL, J. & DOUGAN, G. 2011. Sequence-based analysis uncovers an abundance of non-coding RNA in the total transcriptome of Mycobacterium tuberculosis. *PLoS Pathogens,* 7**,** e1002342.

BELL, K., SEBAIHIA, M., PRITCHARD, L., HOLDEN, M., HYMAN, L., HOLEVA, M., THOMSON, N., BENTLEY, S., CHURCHER, L. & MUNGALL, K. 2004. Genome sequence of the enterobacterial phytopathogen Erwinia carotovora subsp. atroseptica and characterization of virulence factors. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 11105-11110.

BLANKENBERG, D., KUSTER, G. V., CORAOR, N., ANANDA, G., LAZARUS, R., MANGAN, M., NEKRUTENKO, A. & TAYLOR, J. 2010. Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology***,** 19.10. 1-19.10. 21.

BOBROVSKYY, M. & VANDERPOOL, C. K. 2013. Regulation of bacterial metabolism by small RNAs using diverse mechanisms. *Annual Review of Genetics,* 47**,** 209-232.

BURGE, S. W., DAUB, J., EBERHARDT, R., TATE, J., BARQUIST, L., NAWROCKI, E. P., EDDY, S. R., GARDNER, P. P. & BATEMAN, A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Research,* 41**,** D226-D232.

CHAO, Y., PAPENFORT, K., REINHARDT, R., SHARMA, C. M. & VOGEL, J. 2012. An atlas of Hfq-bound transcripts reveals 3′ UTRs as a genomic reservoir of regulatory small RNAs. *The EMBO Journal,* 31**,** 4005-4019.

CHARKOWSKI, A., BLANCO, C., CONDEMINE, G., EXPERT, D., FRANZA, T., HAYES, C., HUGOUVIEUX-COTTE-PATTAT, N., SOLANILLA, E. L., LOW, D. & MOLELEKI, L. 2012. The role of secretion systems and small molecules in soft-rot enterobacteriaceae pathogenicity. *Annual Review of Phytopathology,* 50**,** 425-449.

CHEN, Y., INDURTHI, D. C., JONES, S. W. & PAPOUTSAKIS, E. T. 2011. Small RNAs in the genus Clostridium. *MBio,* 2**,** e00340-10.

CHO, S. H., LEI, R., HENNINGER, T. D. & CONTRERAS, L. M. 2014. Discovery of ethanol responsive small RNAs in Zymomonas mobilis. *Applied and Environmental Microbiology***,** AEM. 00429-14.

CROS, M.-J., DE MONTE, A., MARIETTE, J., BARDOU, P., GRENIER-BOLEY, B., GAUTHERET, D., TOUZET, H. & GASPIN, C. 2011. RNAspace. org: An integrated environment for the prediction, annotation, and analysis of ncRNA. *RNA,* 17**,** 1947-1956.

CROUCHER, N. J., FOOKES, M. C., PERKINS, T. T., TURNER, D. J., MARGUERAT, S. B., KEANE, T., QUAIL, M. A., HE, M., ASSEFA, S. & BÄHLER, J. 2009. A simple method for directional transcriptome sequencing using Illumina technology. *Nucleic Acids Research,* 37**,** e148-e148.

CUI, Y., CHATTERJEE, A. & CHATTERJEE, A. K. 2001. Effects of the two-component system comprising GacA and GacS of Erwinia carotovora subsp. carotovora on the production of global regulatory rsmB RNA, extracellular enzymes, and harpinEcc. *Molecular Plant-Microbe Interactions,* 14**,** 516-526.

FENG, J., MEYER, C. A., WANG, Q., LIU, J. S., LIU, X. S. & ZHANG, Y. 2012. GFOLD: a generalized fold change for ranking differentially expressed genes from RNA-seq data. *Bioinformatics,* 28**,** 2782-2788.

FOZO, E. M., HEMM, M. R. & STORZ, G. 2008. Small toxic proteins and the antisense RNAs that repress them. *Microbiology and Molecular Biology Reviews,* 72**,** 579-589.

GIARDINE, B., RIEMER, C., HARDISON, R. C., BURHANS, R., ELNITSKI, L., SHAH, P., ZHANG, Y., BLANKENBERG, D., ALBERT, I. & TAYLOR, J. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Research,* 15**,** 1451-1455.

GOECKS, J., NEKRUTENKO, A. & TAYLOR, J. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology,* 11**,** R86.

GOGOLEVA, N., SHLYKOVA, L., GORSHKOV, V. Y., DAMINOVA, A. & GOGOLEV, Y. V. 2014. Effect of topology of quorum sensing-related genes in Pectobacterium atrosepticum on their expression. *Molecular Biology,* 48**,** 583-589.

GORSHKOV, V., PETROVA, O., GOGOLEVA, N. & GOGOLEV, Y. 2010. Cell-to-cell communication in the populations of enterobacterium Erwinia carotovora ssp. atroseptica SCRI1043 during adaptation to stress conditions. *FEMS Immunology and Medical Microbiology,* 59**,** 378-385.

GORSHKOV, V. Y., PETROVA, O., MUKHAMETSHINA, N., AGEEVA, M., MULYUKIN, A. & GOGOLEV, Y. V. 2009. Formation of “nonculturable” dormant forms of the phytopathogenic enterobacterium Erwinia carotovora. *Microbiology,* 78**,** 585-592.

GOTTESMAN, S. & STORZ, G. 2011. Bacterial small RNA regulators: versatile roles and rapidly evolving variations. *Cold Spring Harbor Perspectives in Biology,* 3**,** a003798.

GRAVELEY, B. R. 2008. Molecular biology: power sequencing. *Nature,* 453**,** 1197-1198.

GUO, M. S., UPDEGROVE, T. B., GOGOL, E. B., SHABALINA, S. A., GROSS, C. A. & STORZ, G. 2014. MicL, a new σE-dependent sRNA, combats envelope stress by repressing synthesis of Lpp, the major outer membrane lipoprotein. *Genes and Development,* 28**,** 1620.

HENGGE-ARONIS, R. 2002. Signal transduction and regulatory mechanisms involved in control of the σS (RpoS) subunit of RNA polymerase. *Microbiology and Molecular Biology Reviews,* 66**,** 373-395.

HOE, C.-H., RAABE, C. A., ROZHDESTVENSKY, T. S. & TANG, T.-H. 2013. Bacterial sRNAs: regulation in stress. *International Journal of Medical Microbiology,* 303**,** 217-229.

HOFACKER, I. L., FONTANA, W., STADLER, P. F., BONHOEFFER, L. S., TACKER, M. & SCHUSTER, P. 1994. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly,* 125**,** 167-188.

KULKARNI, R. V. & KULKARNI, P. R. 2007. Computational approaches for the discovery of bacterial small RNAs. *Methods,* 43**,** 131-139.

LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods,* 9**,** 357-359.

LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G. & DURBIN, R. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics,* 25**,** 2078-2079.

LIU, Y., CUI, Y., MUKHERJEE, A. & CHATTERJEE, A. K. 1998. Characterization of a novel RNA regulator of Erwinia carotovora ssp. carotovora that controls production of extracellular enzymes and secondary metabolites. *Molecular Microbiology,* 29**,** 219-234.

LIVAK, K. J. & SCHMITTGEN, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2− ΔΔCT method. *Methods,* 25**,** 402-408.

LIVNY, J. 2012. Bioinformatic Discovery of Bacterial Regulatory RNAs Using SIPHT. *Bacterial Regulatory RNA.* Springer.

LIVNY, J. & WALDOR, M. K. 2007. Identification of small RNAs in diverse bacterial species. *Current Opinion in Microbiology,* 10**,** 96-101.

MAJDALANI, N., CHEN, S., MURROW, J., ST JOHN, K. & GOTTESMAN, S. 2001. Regulation of RpoS by a novel small RNA: the characterization of RprA. *Molecular Microbiology,* 39**,** 1382-1394.

MANDIN, P. & GOTTESMAN, S. 2010. Integrating anaerobic/aerobic sensing and the general stress response through the ArcZ small RNA. *The EMBO Journal,* 29**,** 3094-3107.

MIKA, F. & HENGGE, R. 2013. Small regulatory RNAs in the control of motility and biofilm formation in E. coli and Salmonella. *International Journal of Molecular Sciences,* 14**,** 4560-4579.

MIKA, F. & HENGGE, R. 2014. Small RNAs in the control of RpoS, CsgD, and biofilm architecture of Escherichia coli. *RNA Biology,* 11**,** 494-507.

MIRANDA-RÍOS, J., NAVARRO, M. & SOBERÓN, M. 2001. A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proceedings of the National Academy of Sciences,* 98**,** 9736-9741.

MITRA, A., KESARWANI, A. K., PAL, D. & NAGARAJA, V. 2011. WebGeSTer DB—a transcription terminator database. *Nucleic Acids Research,* 39**,** D129-D135.

MIYAKOSHI, M., CHAO, Y. & VOGEL, J. 2015. Regulatory small RNAs from the 3′ regions of bacterial mRNAs. *Current Opinion in Microbiology,* 24**,** 132-139.

MUKHERJEE, A., CUI, Y., LIU, Y., DUMENYO, C. K. & CHATTERJEE, A. K. 1996. Global regulation in Erwinia species by Erwinia carotovora rsmA, a homologue of Escherichia coli csrA: repression of secondary metabolites, pathogenicity and hypersensitive reaction. *Microbiology,* 142**,** 427-434.

NAWROCKI, E. P., BURGE, S. W., BATEMAN, A., DAUB, J., EBERHARDT, R. Y., EDDY, S. R., FLODEN, E. W., GARDNER, P. P., JONES, T. A. & TATE, J. 2014. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research***,** gku1063.

NOÉ, L. & KUCHEROV, G. 2005. YASS: enhancing the sensitivity of DNA similarity search. *Nucleic Acids Research,* 33**,** W540-W543.

NUDLER, E. & MIRONOV, A. S. 2004. The riboswitch control of bacterial metabolism. *Trends in Biochemical Sciences,* 29**,** 11-17.

PEER, A. & MARGALIT, H. 2011. Accessibility and evolutionary conservation mark bacterial small-RNA target-binding regions. *Journal of Bacteriology,* 193**,** 1690-1701.

PÉREZ-MENDOZA, D., COULTHURST, S. J., SANJUÁN, J. & SALMOND, G. P. 2011. N-Acetylglucosamine-dependent biofilm formation in Pectobacterium atrosepticum is cryptic and activated by elevated c-di-GMP levels. *Microbiology,* 157**,** 3340-3348.

PERKINS, T. T., KINGSLEY, R. A., FOOKES, M. C., GARDNER, P. P., JAMES, K. D., YU, L., ASSEFA, S. A., HE, M., CROUCHER, N. J. & PICKARD, D. J. 2009. A strand-specific RNA–Seq analysis of the transcriptome of the typhoid bacillus salmonella typhi. *PLoS Genetics,* 5**,** e1000569.

PETROVA, O., GORSHKOV, V., DAMINOVA, A., AGEEVA, M., MOLELEKI, L. & GOGOLEV, Y. 2013. Stress response in Pectobacterium atrosepticum SCRI1043 under starvation conditions: adaptive reactions at a low population density. *Research in Microbiology,* 165**,** 119-127.

PETROVA, O., GORSHKOV, V., DAMINOVA, A., AGEEVA, M., MOLELEKI, L. N. & GOGOLEV, Y. 2014. Stress response in Pectobacterium atrosepticum SCRI1043 under starvation conditions: adaptive reactions at a low population density. *Research in Microbiology,* 165**,** 119-127.

PICHON, C., DU MERLE, L., CALIOT, M. E., TRIEU-CUOT, P. & LE BOUGUÉNEC, C. 2012. An in silico model for identification of small RNAs in whole bacterial genomes: characterization of antisense RNAs in pathogenic Escherichia coli and Streptococcus agalactiae strains. *Nucleic Acids Research,* 40**,** 2846-2861.

REPOILA, F., MAJDALANI, N. & GOTTESMAN, S. 2003. Small non-coding RNAs, co-ordinators of adaptation processes in Escherichia coli: the RpoS paradigm. *Molecular Microbiology,* 48**,** 855-861.

RICHARDS, G. R. & VANDERPOOL, C. K. 2011. Molecular call and response: the physiology of bacterial small RNAs. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms,* 1809**,** 525-531.

ROBINSON, M. D., MCCARTHY, D. J. & SMYTH, G. K. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics,* 26**,** 139-140.

RUTHERFORD, K., PARKHILL, J., CROOK, J., HORSNELL, T., RICE, P., RAJANDREAM, M.-A. & BARRELL, B. 2000. Artemis: sequence visualization and annotation. *Bioinformatics,* 16**,** 944-945.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular cloning*, *Cold Spring Harbor Laboratory Press New York*.

SOUTOURINA, O. A., MONOT, M., BOUDRY, P., SAUJET, L., PICHON, C., SISMEIRO, O., SEMENOVA, E., SEVERINOV, K., LE BOUGUENEC, C. & COPPÉE, J.-Y. 2013. Genome-Wide Identification of Regulatory RNAs in the Human Pathogen Clostridium difficile. *PLoS Genetics,* 9**,** e1003493.

URBAN, J. H. & VOGEL, J. 2008. Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. *PLoS Biology,* 6**,** e64.

VOCKENHUBER, M.-P., SHARMA, C. M., STATT, M. G., SCHMIDT, D., XU, Z., DIETRICH, S., LIESEGANG, H., MATHEWS, D. H. & SUESS, B. 2011. Deep sequencing-based identification of small non-coding RNAs in Streptomyces coelicolor. *RNA Biology,* 8**,** 468.

WATERS, L. S. & STORZ, G. 2009. Regulatory RNAs in bacteria. *Cell,* 136**,** 615-628.

WILMS, I., OVERLÖPER, A., NOWROUSIAN, M., SHARMA, C. M. & NARBERHAUS, F. 2012. Deep sequencing uncovers numerous small RNAs on all four replicons of the plant pathogen Agrobacterium tumefaciens. *RNA Biology,* 9**,** 446.

Table 4.1: 3` UTR encoded sRNAs

		sRNA			RPKMs		Expression	Predicted	sRNA promoter and start site			Transcription Factor
sRNA name	**Parent mRNA**	**Start**	**end**	**length**	**sRNA**	**mRNA**	**(based on RPKMs)**	**promoters**	**-35**	**-10**	**TSS**	**binding site**
fwd_4	*rbsB*	14355	14537	183	596.8	102.8	Independent	1	14317	14342	14357	rpoD17, cynR, rpoD15, rpoD16, phoB
fwd_6	*polA*	28634	28755	122	549.8	110.5	Independent	1	28656	28676	28691	cytR, arcA, crp, rpoD15, rpoD17
fwd_15	*ECA0044*	55039	55291	253	66.8	27.4	Independent	1	55009	55029	55044	rpoD17, fis, rpoD15, rpoD16, phoB
fwd_19	*expI*	126355	126501	147	338.5	588.1	Co-expression	1	126418	126435	126450	metR, rpoD16
reg_seq13	*aldA*	139913	140154	242	136.8	30.4	Independent	1	140046	140063	140078	
reg_seq27	*ECA0332*	380584	380826	243	1660.5	389.3	Independent	1	380610	380631	380646	metJ
reg_seq31	*ECA0449*	515673	515910	238	363.3	528.7	Co-expression	1	515790	515813	515828	glpR, ihf, argR2, nagC, argR2, fnr, fis
reg_seq43	*mdH*	758603	758857	255	1759.1	1895.7	Co-expression	2	758751	758770	758789	glpR, fis, arcA, purR
									758444	758464	758479	
reg_seq142	*ECA2516*	2832294	2832530	237	514.2	77.5	Independent	1	2832185	2832205	2832220	purr, rpoD16
fwd_rfam4	*rpiL*	255260	255321	62	775.5	1856.1	Independent					lrp, hns
comp_seq5	*glnA*	34992	35234	243	858.8	1209.3	Co-expression	1	35262	35241	35226	rpoD15, rpoD16, phoB
comp_seq11	*slmA*	164026	164334	309	196.4	454.0	Independent	1	164395	164375	164360	arcA, rpoD17, rpoD15, rpoD16, phoB
comp_seq130	*ECA2950*	3295111	3295347	237	197.9	222.9	Co-expression	2	3295196	3295176	3295161	rpoD16, argR, arcA, ihf
									3295506	3295485	3295470	
rev_rfam22	*glpC*	4651380	4651485	109	348.2	40.8	Independent	1	4651583	4651562	4651547	Ada, rpoE, tyrR, fur, fur
dapZ	*ECA3872 (dapB)*	4332178	4332288	111	20.9	110.8	Independent	1	4332302	4332284	4332269	Ihf, argR2, rpoD16, argR,, fis, crp

Table 4.2: *In silico* predicted sRNA candidates confirmed by strand-specific RNA-seq

sRNA candidate	sRNA class	sRNA type	sRNA start	sRNA end	sRNA length
reg_seq3	3`UTR: *polA*		28634	28755	122
reg_seq3b	IGR	spot42 sRNA	28756	28882	127
reg_seq13	3` UTR: *aldA*		139913	140154	242
reg_seq27	3` UTR: *ECA0332*	TPP riboswitch	380584	380826	243
reg_seq27b	5` UTR: *icc*	isrH (Hfq binding sRNA)	380825	380917	93
reg_seq31	3` UTR: *ECA0449*		515673	515910	238
reg_seq34	3` UTR: *topB*	STAXI sRNA	601635	601862	228
reg_seq34b	antisense: *ECA0527*	STAXI sRNA	601860	602087	228
reg_seq43	3` UTR: *mdh*	Glycine riboswitch	758603	758857	255
reg_seq67	IGR		1185902	1186147	246
reg_seq70	antisense: *ECA1096*		1225748	1226182	435
reg_seq76	5` UTR: *mend*	TPP riboswitch	1379050	1379349	300
reg_seq109	antisense: *osmB*	TPP/ isrH	2217586	2217946	361
reg_seq129	IGR	Trp leader	2602697	2602931	235
reg_seq133	5` UTR: *ansA*	RtT and TPP	2651356	2651598	243
reg_seq142	3` UTR: *ECA2516*		2832294	2832530	237
comp_seq5	3` UTR: *glnL*	TPP/ isrH	34992	35234	243
comp_seq11	3` UTR: *slmA*	isrH	164026	164334	309
comp_seq16	5`UTR: *ECA0353*	TPP	403257	403655	399
comp_seq49	IGR	TPP	1218529	1218729	201
comp_seq55	5` UTR: *ECA1196*	TPP/ isrH	1358123	1358268	146
comp_seq111	IGR	TPP/ isrH	2881355	2881546	192
comp_seq130	3` UTR: *ECA2950*	TPP and RtT	3295111	3295347	237
comp_seq204	5` UTR: *sotB*		4828158	4828349	192
comp_seq217	5` UTR: *ECA4506*		5046219	5046427	209

Table 4.3: Functional annotation of the 68 true (and/or known) sRNAs identified by strand-specific RNA-seq

sRNA name	sRNA type	Characterized in this study (RNAspace)	Previously characterized (Rfam)	Total
RNaseP	Ribozyme		1	1
tmRNA			1	1
***Cis*-regulators**				
	Riboswitches			
TPP riboswitch		6	3	9
TPP or isrH		8		8
glycine riboswitch		2	2	4
FMN			1	1
lysine riboswitch		1		1
MOCO_RNA_motif			1	1
yybP-ykoY			1	1
	RNA elements			
alpha_RBS			1	1
cspA			2	2
greA			1	1
his_leader			1	1
JUMPStart			2	2
leucine operon leader			1	1
P26			1	1
rne5			1	1
RtT		1	2	3
trp_leader			1	1
***trans*-encoded sRNA**				
	sRNA			
STAXI			3	3
6S			1	1
crcB			1	1
csrB			1	1
glmY (*tke1*)			1	1
Rye		1		1
sraC (*ryeA*)			1	1
STnc240			1	1
StyR-44			5	5
t44			1	1
	Hfq binding sRNA			
frnS			1	1
isrH		2		2
glmZ (*sraJ*)			1	1
omrA			1	1
rprA			1	1
ryhB			1	1

ryeB (sdsR)			1	1
sgrS			1	1
spot 42			1	1
antisenseRNA (asRNA)	**asRNA**			
HPnc0260			1	1
Total		21	47	68

Table 4.4: Novel sRNA candidates obtained using conservation analysis

sRNA Name	Strand	Length	sRNA Class
rev_11	-	420	asRNA: ECA0328
rev_13	-	354	asRNA: ECA0388
rev_24	-	489	asRNA: *rcsC*
rev_39	-	300	5` UTR: *ilvG*
rev_41	-	480	IGR/ 5` UTR: *bcsB*
fwd_6	+	122	3` UTR: *ECA3097*
fwd_42	+	480	5` UTR: *zipA*
fwd_44	+	336	asRNA: *ECA0910*
fwd_72	+	426	asRNA: *gudP*

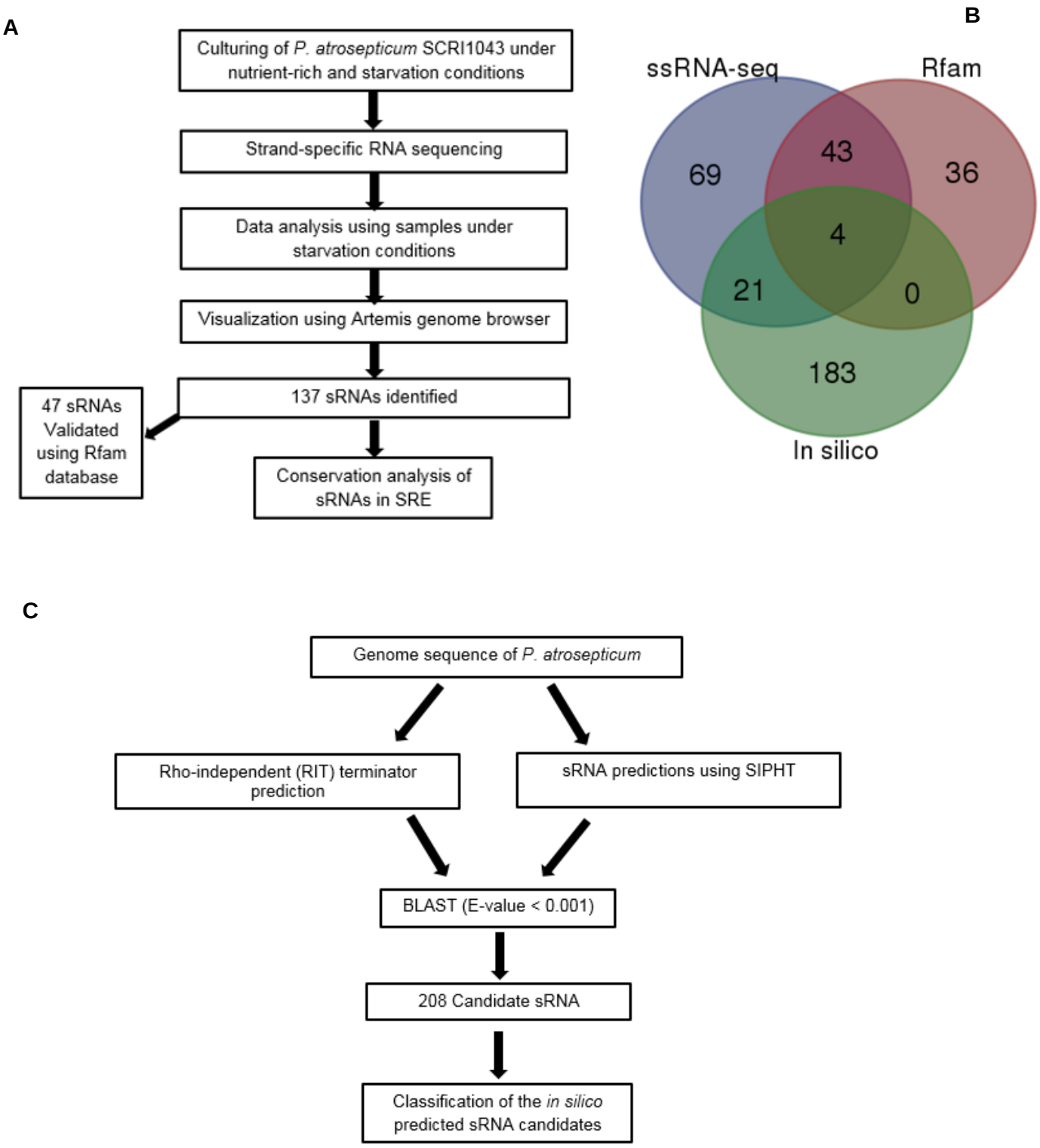


Fig. 4.1: Scheme for sRNA identification. A. Determination of sRNA using strand-specific RNA seq of P. atrosepticum cultured under starvation conditions. B. Comparison of sRNAs identified by strand-specific RNA-seq with sRNA candidates predicted for P. atrosepticum in Rfam database and sRNAs predicted computationally in this study. C. Computational (in silico) sRNA prediction.

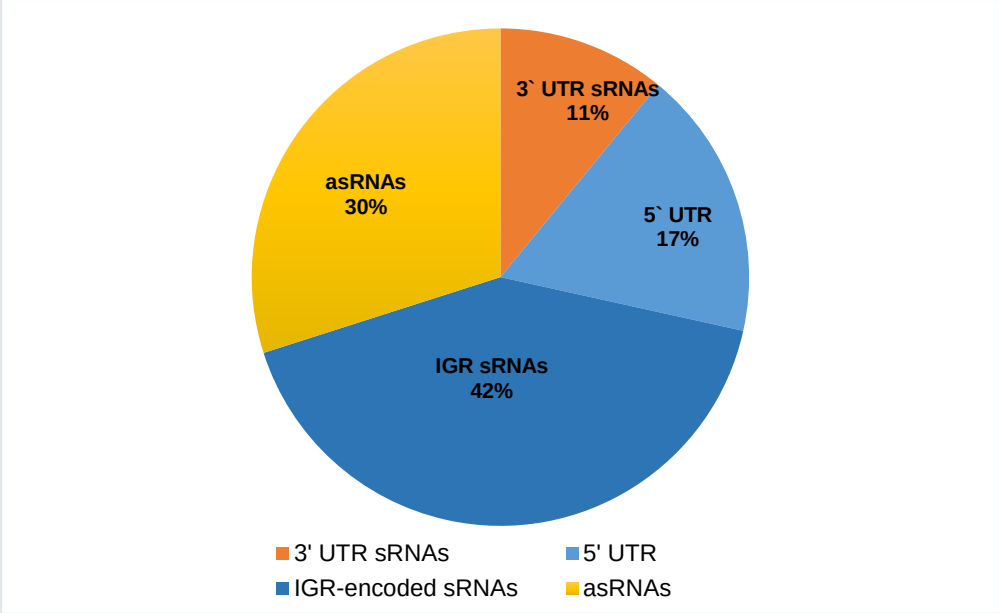


Fig. 4.2: Pie chart showing classification of sRNAs identified using ssRNA-seq into four classes that is IGR/ trans-encoded sRNAs, asRNA, 5`UTR (riboswitches), and 3` UTR sRNAs, based on their proximity and location with regards to CDS regions.

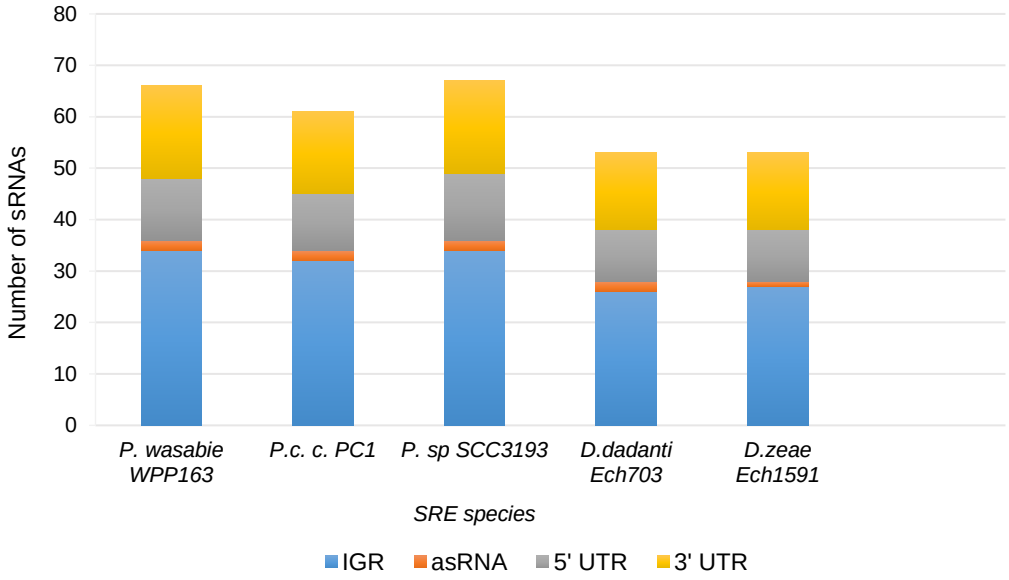


Fig. 4.3: Summary of conserved sRNAs in soft rot enterobacterieace

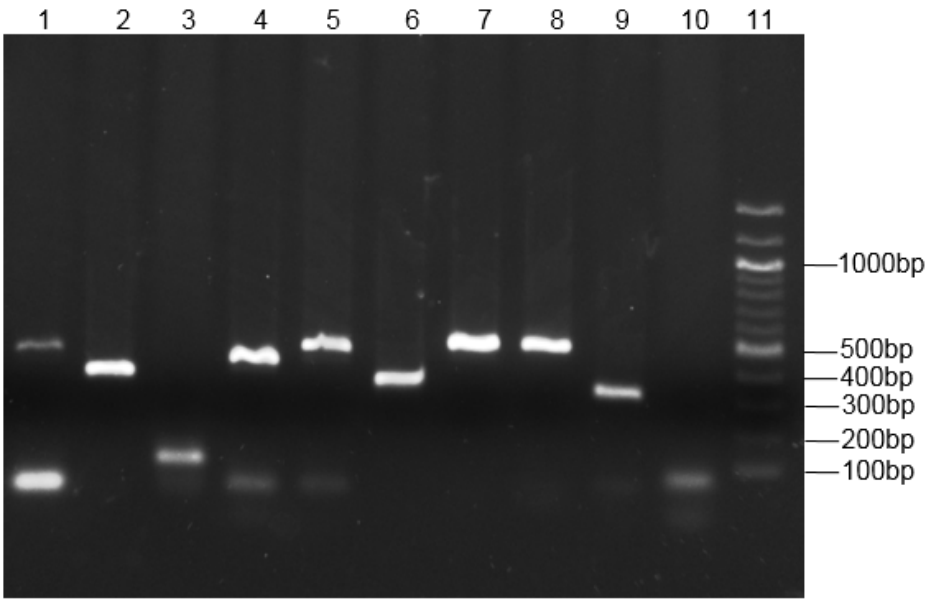


Fig. 4.4: Validation of novel sRNA expression by RT-PCR: Agarose gel electrophoresis of the PCR amplicon fragments of the 9 novel sRNAs. Lane 1. rev_41, Lane 2. rev_13, Lane 3. fwd_6, Lane 4. rev_11, Lane 5. fwd_72, Lane 6. fwd_44, Lane 7. fwd_42, Lane 8. rev_24, Lane 9. rev_39, Lane 10. No reverse transcriptase control, Lane 11. 100 bp DNA Ladder.

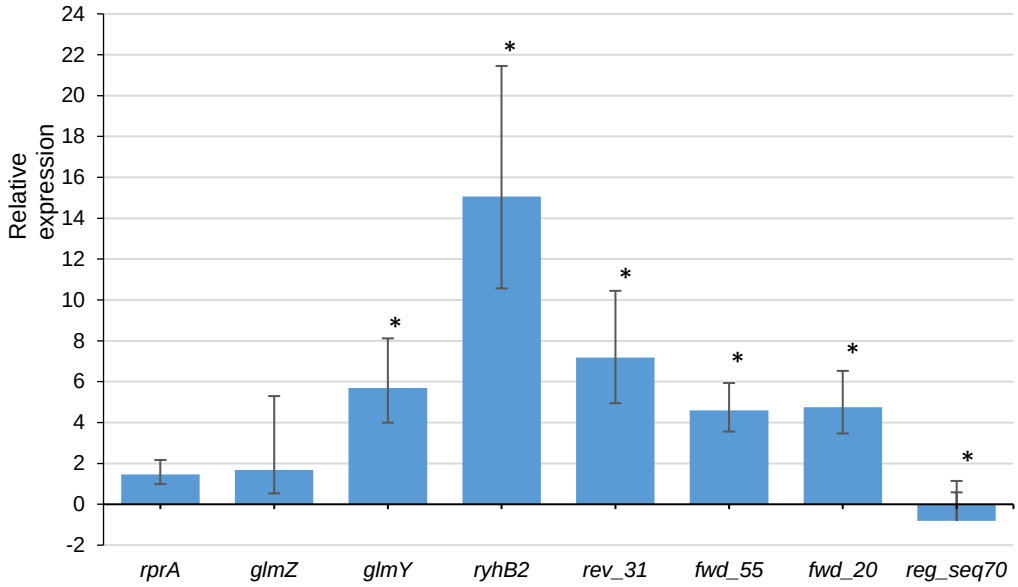


Fig. 4.5: RT-qPCR validation of RNA-seq expression analysis. Relative expression changes of sRNAs were determined using the $2^{-\Delta\Delta ct}$ method by comparing expression in starvation conditions to nutrient-rich. Error bars indicate the standard error of three independent biological replicates. Asterisks represent significant difference at $p<0.05$ (Students t-test).

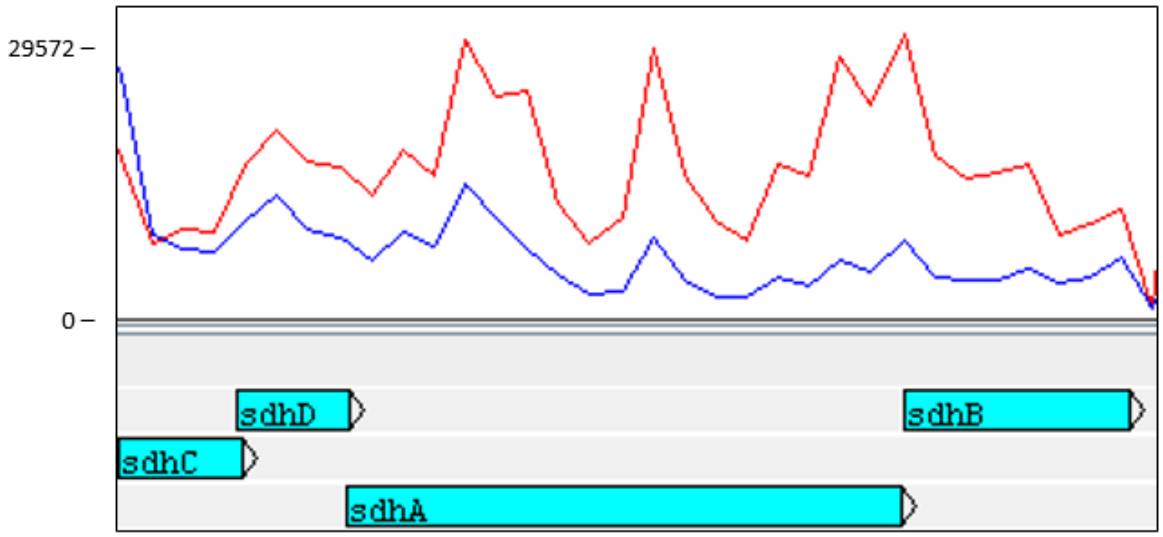


Fig. 4.6: The expression of the sdhCDAB operon is relatively lower under starvation compared to growth promoting conditions. Reads mapped from the nutrient-rich condition are represented by the red line. The blue line represents mapped reads from the starvation conditions. Annotated features are labelled below the plot in blue blocks. The y-axis shows the read coverage per coding region (CDS).

CHAPTER FIVE

5.0 Concluding Summary

Pectobacterium carotovorum subsp. *brasiliense* (*Pcb*) is a member of the soft rot *Enterobacteriaceae* (SRE) family that causes tuber soft rot and blackleg diseases of stems in potato, its main host. Currently, no chemical strategies exist to control infection by members of the SRE once disease has been established in the plants. *Pcb* is a broad host range aggressive soft rot pectobacterium and is emerging as a phytopathogen of global significance with the potential to cause severe yield losses in the potato industry. *Pcb* often migrates from infected mother tubers and colonizes and spreads via the vascular systems of stems, mostly resulting in blackleg disease. Therefore, an understanding of the inducible defense responses in stems of potato is important, in order to determine the molecular basis of resistance to soft rot pathogens exhibited by some potato cultivars. Currently, no transcriptome-wide studies have focused on the inducible defense responses in potato stems during infection by SRE phytopathogens, particularly *Pcb*. Therefore, the main objective of this study was to unravel the global potato defense transcriptome against *P. c. brasiliense* 1692 (*Pcb1692*) and determine the time-course of defense response activation.

In this work, we employed a time-course RNA-seq analysis and strand-specific RNA-seq to investigate potato global defense responses during infection by *Pcb1692*. Comparison of expressed genes between a susceptible potato cultivar (*Solanum tubersoum* cv Valor) and a tolerant cultivar (*S. tuberosum* cv BP1) undertaken in Chapter 2 revealed a total number of 8300, 7393, 12428, 6127, and 8499 differentially expressed genes, identified at 0, 6, 12, 24 and 72 hours post inoculation (hpi) within the time-course, respectively. The highest number of differentially expressed genes was observed at 12 hpi, suggesting that stem-based defense responses are activated early in the tolerant cultivar, influencing plant immunity at later infection stages. The identified differentially expressed genes were involved in pathogen recognition and MAPK signal transduction; in addition several transcription factors were

identified as well as genes involved in ethylene biosynthesis and signaling pathway. Collectively, expression profiles of the differentially expressed genes and gene ontology enrichment analysis revealed that the MPK3/MPK6 and MPK4 genes, and WRKY-type transcription factors (e.g., AtWRKY33) are key defense genes in potato immunity. These genes were only up-regulated in the tolerant cultivar indicating that they are probable key components in the potato defense mechanisms against *Pcb1692* infection. Furthermore, several secondary wall biosynthetic genes important in secondary cell wall modifications including MYB83, *SND1* and *SND2* were significantly up-regulated in the tolerant cultivar. Taken together, our results show that immunity of potato plants to soft rot pectobacteria, especially *Pcb*, is quantitative, requiring many genes to confer resistance. Therefore, we propose a potato defense model against the necrotrophic *Pcb1692* which is centered on Pattern-Triggered Immunity, including pathogen perception, MAPK cascade signaling, ethylene hormonal signaling and activation of downstream plant defenses such as cell wall reinforcements (Fig. 5.1). As neither chemical treatments nor effective *R* gene resistance breeding strategies are available for the control of soft rot pathogens, information generated in Chapter 2 of this study becomes pertinent in breeding for resistance programs focused on quantitative resistance. Furthermore, we envision that information generated in Chapter 2 will be useful to the research community and provide a platform for additional functional studies for confirmation of potato defense genes.

In addition, by using the strand-specific RNA-seq data, we identified 1113 potato candidate long intergenic noncoding RNAs (lincRNAs) expressed in the two potato cultivars in Chapter 3. Therefore, it was important to determine the set of lincRNAs responsive to *Pcb1692* infection in potato as this will provide valuable insights into our current understanding of the global potato defense transcriptome and various defense-related regulatory mechanisms. To date, no lincRNAs have been implicated in potato defense mechanisms. Thus, we hypothesized that lincRNAs could be involved in potato defence mechanisms and therefore differentially expressed in the tolerant cultivar compared to the susceptible cultivar.

Consequently, we identified 559 novel potato lincRNA candidates that showed significant differential expression in both cultivars compared to mock-inoculated controls after infection by *Pcb1692*. Furthermore, co-expression analysis associated 17 of these lincRNAs with 12 potato defense-related genes. These results suggest that lincRNAs have potential functional roles in potato defense regulation. In addition, the work outlined in Chapter 3 also provided the first library of potato lincRNAs implicated in defense responses. However, further functional studies will be necessary in the future to experimentally validate the roles played by the 17 lincRNAs co-expressed with defense-related genes in potato stem-based immune responses. Additional future work will involve determining the interaction of these lincRNAs and components of pattern-triggered immunity in potato immune responses against necrotrophic pathogens.

Furthermore, in order to gain deeper knowledge about regulatory mechanisms involved in survival strategies of *Pectobacterium* species under nutrient-limiting conditions such as xylem vessels during stem colonization, we analyzed starvation-related small RNAs (sRNAs) in *Pectobacterium atrosepticum*. The regulatory roles of sRNAs in bacterial adaptation to nutrient limited conditions have previously been reported in other species such as *Escherichia coli* and *Salmonella*. Herein, our work revealed that small RNAs (sRNAs) are actively expressed in *P. atrosepticum* under nutrient limiting conditions. Specifically, we identified 137 sRNAs, and among these, 68 were differentially expressed between nutrient-rich and starvation conditions *in vitro*. Since many of the identified sRNAs were starvation-induced, the results of our study (in Chapter 4) support the notion that sRNAs play key roles in bacterial adaptive response. Additionally, because SRE phytopathogens colonize nutrient-deficient xylem vessels in potato stems during infection, it is possible that sRNAs are extensively involved in the adaptation and survival of *Pectobacterium species* in the stem vasculature. This knowledge is important in understanding the host-pathogen interaction between *Pectobacterium* and potato, especially in terms of *Pectobacterium* pathogenicity and survival mechanisms *in planta*. In the future, it

is necessary to determine and experimentally confirm mRNA targets of these starvation-related sRNAs, via post-transcriptional reporter fusions of target mRNAs.

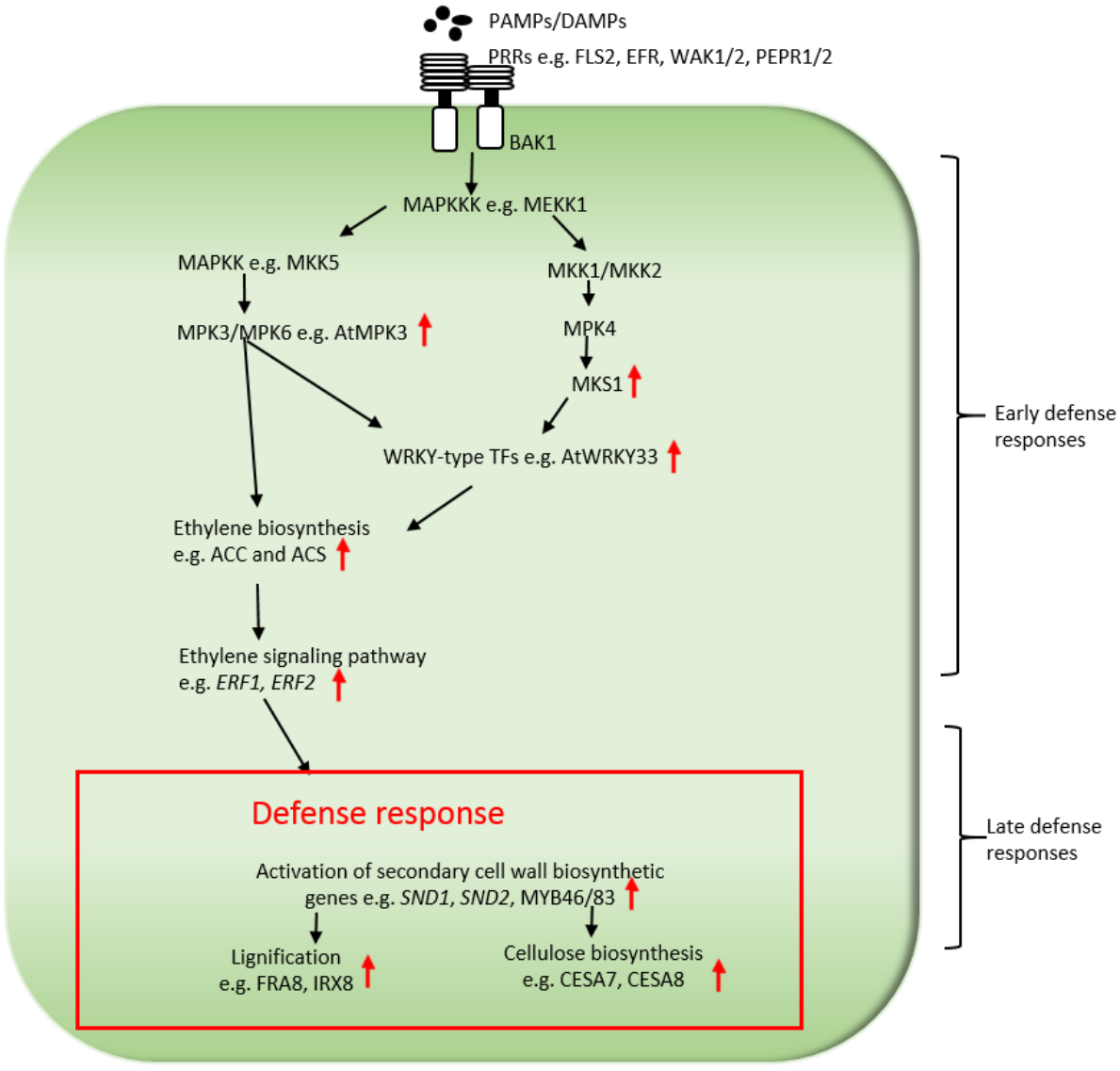


Fig. 5.1. Components of pattern-triggered immunity regulating defense responses against the necrotrophic *P. carotovorum* subsp. *brasiliense* infection in a tolerant potato cultivar (*S. tuberosum* cv. BP1). In Arabidopsis, MAPK cascades, ethylene signaling pathway and WRKY33 have been shown as key immune components against necrotrophs in plant immunity. Red arrows show up-regulation only in tolerant cultivar compared to susceptible cultivar. Red box shows some of the activated downstream defense genes.