Identification and characterization of late blight resistance genes in potato

Sripad Joshi

Department of Plant Science

McGill University, Montreal

August 2020

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Doctor of Philosophy

Dedicated to my grandmother

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ACT	Agmatine Coumaroyl transferase
AGL	Agamous-like protein
AME	Accurate mass error
APG	Angiosperm Phylogeny Group
AUDPC	Area under disease progress curve
bHLH	basic helix-loop-helix
BLAST	Basic local alignment search tool
Bp	Base pair
bZIP	Basic Leucine Zipper Domain
CAGE	Cap analysis of gene expression
CaMV	Cauliflower mosaic virus
Cas	CRISPR associated protein
cDNA	Complementary deoxyribonucleic acid
CDPK	Calcium-dependent protein kinase
CDS	CoDing sequence
ChIP	Chromatin immunoprecipitation
CoA	Co Enzyme A
CRISPR	Clustered regularly interspaced short palindromic repeats
CTAB	Cetyl trimethylammonium bromide
Da	Dalton
DEGs	Differentially expressed genes
DFI	direct flow injection
DNA	Deoxyribonucleic acid
Dof	DNA-binding with one finger
Dpi	Days post inoculation
DSB	Double strand break
ENCODE	Encyclopedia of DNA Elements
ERF	Ethylene response factor
ESI	Electrospray ionization

ETI	Effector triggered immunity
FC	Fold change
FPKM	Fragments Per Kilobase of transcript per Million mapped reads
FT	Feruloyl transferase
FTMS	Fourier transform-ion cyclotron resonance mass spectrometry
GFP	Green fluorescent protein
gRNA	Guide ribonucleic acid
HCAAs	Hydroxycinnamic acid amides
HDR	Homology directed repair
HHMER	Hidden Markov models
Hpi	Hours post inoculation
HRAM	High-resolution, accurate-mass
HSF	Heat shock factor
Kcal/mol	Kilo calories per mole
KEGG	Kyoto Encyclopedia of Genes and Genomes
LB	Late blight
LC-HRMS	Liquid chromatography high resolution mass spectrometry
LC-MS	Liquid chromatography mass spectrometry
MALDI	Matrix-assisted laser desorption/ionization
MAPK	Mitogen activated protein kinase
MNs	Mega nucleases
mRNA	Messenger ribonucleic acid
MS medium	Murashige and Skoog medium
MSA	Multiple sequence alignment
MYB	Myeloblastosis
NBS-LRR	Nucleotide-binding site leucine-rich repeat
NCBI	National Center for Biotechnology Information
NHEJ	Non homologous end joining
NMR	Nuclear Magnetic Resonance
PAMP	Pathogen-associated molecular pattern
PCR	Polymerase chain reaction

PDA	Potato Dextrose Agar
PDB	Protein data bank
PGSC	Potato Genome Sequencing Consortium
Ppm	Parts per million
PR	Pathogenesis related
qRT-PCR	Quantitative reverse transcription PCR
QTL	Quantitative trait loci
RAP	RNA polymerase II-associated protein
RM	Resistant mock inoculated
RMSD	Root-mean-square deviation
RP	Resistant pathogen inoculated
RR	Resistance Related
RRC	Resistance related constitutive
RRI	Resistance related induced
SM	Susceptible mock inoculated
SNP	Single nucleotide polymorphism
SP	Susceptible pathogen inoculated
StACT	*Solanum tuberosum* Agmatine coumaroyl transferase
StFT	*Solanum tuberosum* Feruloyl transferase
T-DNA	Transfer deoxy ribonucleic acid
TFs	Transcription factors
TOF	Time of flight
ZFN	Zinc finger nucleases

Abstract

Potato (*Solanum tuberosum)* is one of the major non-cereal staple food crops grown across the world. Canada produces 4.4 million tonnes of potato. The major constraints in potato production are the biotic and abiotic stresses. Among the biotic stresses, late blight of potato is one of the destructive diseases, caused by an oomycete, *Phytophthora infestans*. Late blight of potato causes up to 40% yield loss worldwide. Resistance to late blight is either qualitative or quantitative. Even though quantitative resistance is durable, the genetic bed rock underlying resistance is not well deciphered, which limits its applications. The objective of this study was to identify the resistance genes and their mechanisms, in a resistant genotype (Libertas) and a commercial susceptible genotype (AG704). The first study aimed at transcriptomics and metabolomics to identify the induced resistance related metabolites and genes, including the large impact transcription factors which regulate these genes and metabolites that are in higher fold change in a resistant genotype than in a susceptible genotype. A total of 160 induced metabolites and 611 induced genes were identified. A Transcription Factor (TF) enrichment study identified a total of 134 regulatory TFs, which were highly enriched in the promoters of induced genes. A correlation study among the induced genes further mapped several primary and secondary TFs. Among these *bHLH66, MYB61, NAC56, WRKY51, MYB like, ERF RAP2-3* and *MADS-box AGL15* had regulating sites in more than 208 downstream genes, of which many were secondary TFs. Hence, a two-tier transcriptional regulation of defense response genes was mapped. The metabolic profiling identified Hydroxycinnamic acid amides (HCAAs). Feruloylagmatine was highly accumulated in the resistant genotype post pathogen inoculation. Two major genes *StACT1* and *StACT2* which might be responsible were characterized by gene sequencing, qPCR and *in silico* protein docking data and it revealed the *StACT2* gene to be mainly responsible for feruloylagmatine accumulation in potato. We found a *NAC* TF regulating *StACT2* promoter region and it was identified as *NAC72,* which was accumulated more in the resistant than in the susceptible genotype. When the *NAC72* CDS and promoter regions were sequenced a mutation was found in the promoter of *NAC72.* This gene was edited in a late blight susceptible commercial genotype, AG704, to enhance resistance. The expressions of genes *StACT2* and *NAC72* were significantly increased, however, the disease severity showed no significant difference. It is possible that there may be a missing chain in the hierarchy of genes involved in the regulation

and/or in the metabolic pathway network to synthesize the metabolite feruloylagmatine, which is known to enhance disease resistance through deposition of these metabolites to reinforce the secondary cell walls, thus containing the pathogen to initial infection area. Discovery and editing of other mutated gene(s) should enhance feruloylagmatine production, as in resistant genotype. The other genes identified here also can be used in future for genome editing to increase RR metabolite accumulation in potato to enhance late blight and other disease resistance in potato.

Résumé

La pomme de terre (*Solanum tuberosum*) est l'une des principales cultures vivrières de base non céréalières cultivées dans le monde. Le Canada produit 4,4 million de tonnes de pommes de terre. Les contraintes majeures dans la production de pommes de terre sont les stress biotiques et abiotiques. Parmi les stress biotiques, le mildiou de la pomme de terre est l'une des maladies les plus destructrices, causée par un oomycète, *Phytophthora infestans*. Le mildiou de la pomme de terre cause jusqu'à 40% de perte de rendement dans le monde. La résistance au mildiou est soit qualitative, soit quantitative. Même si la résistance quantitative est durable, la structure génétique sous-jacente à la résistance n'est pas bien déchiffrée, ce qui limite ses applications. L'objectif de cette étude était d'identifier les gènes de résistance et leurs mécanismes dans un génotype résistant (Libertas) et un génotype commercial sensible (AG704). La première étude était basée sur la transcriptomique et la métabolomique pour identifier les métabolites et les gènes liés à la résistance, y compris les facteurs de transcription clés qui régulent ces gènes et métabolites et qui sont plus fortement modifiés dans un génotype résistant que dans un génotype sensible. Un total de 160 métabolites induits et 611 gènes induits ont été identifiés. Une étude d'enrichissement en facteur de transcription (TF) a identifié un total de 134 TF régulateurs qui étaient hautement enrichis en promoteurs de gènes induits. Une étude de corrélation entre les gènes induits a également cartographié plusieurs TF primaires et secondaires. Parmi ceux-ci *bHLH66, MYB61, NAC56, WRKY51, MYB like, ERF RAP2-3* et *MADS-box AGL15* avaient des sites de régulation dans plus de 208 gènes en aval, dont beaucoup étaient des TFs secondaires. Par conséquent, une régulation transcriptionnelle à deux niveaux des gènes de réponse de défense a été cartographiée. Le profilage métabolique a identifié les amides d'acide hydroxycinnamique (HCAA). La feruloylagmatine était fortement accumulée dans le génotype résistant après l'inoculation du pathogène. Deux gènes majeurs *StACT1* et *StACT2* pouvant être responsables de cette accumulation ont été caractérisés par séquençage des gènes, PCR quantitatif et protéines d'ancrage *in silico*, et il a révélé que le gène *StACT2* était principalement responsable de l'accumulation de feruloylagmatine dans la pomme de terre. Nous avons trouvé une région du promoteur *StACT2* régulant le TF *NAC*. Celui-ci a été identifié comme *NAC72* et s'est accumulé davantage dans le génotype résistant que dans le génotype sensible. Lorsque les régions CDS et du promoteur de *NAC72* ont été séquencées, une mutation a été trouvée dans le promoteur de *NAC72*. Ce gène a été

modifié dans un génotype commercial sensible au mildiou, AG704, pour augmenter la résistance. Dans AG704, les expressions des gènes *StACT2* et *NAC72* étaient significativement augmentées; cependant, la gravité de la maladie n'a montré aucune différence significative. Il est possible qu'il y ait un maillon manquant dans la hiérarchie des gènes impliqués dans la régulation et/ou dans le réseau des voies métaboliques pour synthétiser le métabolite feruloylagmatine, ce dernier étant connu pour améliorer la résistance aux maladies en se déposant afin de renforcer les parois cellulaires secondaires et contenir ainsi l'agent pathogène dans la zone d'infection initiale. La découverte et l'édition d'autres gènes mutés devraient améliorer la feruloylagmatine, comme dans le génotype résistant. Les autres gènes identifiés ici peuvent également être utilisés à l'avenir pour l'édition du génome de la pomme de terre afin de favoriser l'accumulation de métabolites RR et ainsi améliorer la résistance au mildiou et à d'autres maladies.

Acknowledgements

I have been very lucky to receive help and advice from several people during the course of my study at McGill University, and I cannot write each and every name here, but I am in debt for all your help and collaboration. I am extremely grateful and privileged to seek guidance from my beloved supervisor **Prof. AJJAMADA C. KUSHALAPPA**. I am also grateful for his critical comments, constant support and valuable advice time and again, during the entire course of my studies. His balanced way of dealing with things through the course of my investigation has helped me work diligently. It was pleasure and privilege, to be associated with him and his lab during this course of study. Besides my supervisor, I would like to thank the rest of my thesis committee: **Dr. JEAN-BENOIT CHARRON** and **Dr. SÉBASTIEN P. FAUCHER**, for their constant help in all aspects, criticism and insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives. I should specially thank Dr. CHARRON for his helpful suggestions in carrying out this research work, providing access to his lab and guidance as the graduate program director.

I must express heartfelt thanks to **Dr. MARTINA STROMVIK** who is the chair of Plant Science department, for her tremendous support and being one of my course supervisors. It gives me immense pleasure to express my sincere gratitude and heartfelt thanks to all professors of Plant Science faculty. I must also thank the plant science staff, Mr. Guy Rimmer, Mr. Ian Ritchie, Ms. Carolyn Bowes and Mrs. Lynn Bachand and Ms. Erin Minnett for their cooperation and timely help to carry out greenhouse experiments and administrative tasks successfully. My sincere thanks also goes to **Dr. OLIVIA WILKINS**, **Dr. DANIELLE DONNELLY**, and **Dr. ROBIN BEECH**, for serving as members of my comprehensive exam committee. I want to convey my special gratitude to **Dr. RAJ DUGGAVATHI** for providing constant support by giving timely advice and motivation and also providing help in terms of lab equipment's for conducting qRT-PCR, without which my research would not finish. I also thank Ms. SUSAN GREGUS for assisting in administrative activities and helping in timely disbursement of scholarships.

I thank my fellow lab mates Ms. Nancy Soni, Dr. Niranjan Hegde, Mr. Russiachand Singh Heikham, Mr. Bikram Poudel, Mr. Krishna Kumar and Ms. Chaitra Nagaraju, for supporting my work and stimulating discussions, for the sleepless nights we were working together finding

solutions to problems, and for all the fun we have had in the last few years. I also extend my sincere gratitude towards the visiting Professors we had in lab, Prof. Farhad Nazarian-Firouzabadi and Dr. Huali Xue, for their collaboration and help during my work. I must also thank the interns whom I could train during their visit, Mrs. Roshni Nair and Ms. Kiana Khazaei, who also helped me immensely in my work. I wish to express my sincere gratitude towards senior lab members, Dr. Yogendra Kalenahalli, Dr. Shailesh Karre, Dr. Uday Kage, Dr. Dhananjay Dhokane and Dr. Shivappa Hukkeri from whom I could learn a lot during the initial days of my studies, and they have helped me a lot to start my work. I am grateful to all those with whom I have had the pleasure to work and learn, especially all my colleagues and friends in the plant science department and friends at the Macdonald Campus.

It would not have been possible to achieve anything without the support of my family. I wish to express my gratitude towards my beloved parents **Mr. SHIRISH JOSHI** and **Mrs. SARITA JOSHI,** who have helped me grow as a person and made me what I am today. It would not have been possible to start or complete anything without the constant support of my loving wife **Mrs. KSHITIJA PAWAR**, who has always been there for me, in ups and down, and motivated me at my worst times. I also thank my younger sister **Ms. SAMIKSHA JOSHI** for her kinds words and support whenever it was needed.

This work would not have been possible without the financial support from all the funding agencies. I must first thank **MR. ANDRE GAGNON, AND** Progest 2001 Inc., who allowed us to work on the potato genotype they developed and also for the financial support they provided. I must thank **MITACS** for the financial support as ACCLERATE internships. I also thank **MCGILL UNIVERSITY** for the GREAT award through which I could attend conferences and expand my knowledge, for graduate excellence award and research assistantship and also for the Frank P Jones Bursary. I also thank **Natural Sciences and Engineering Research Council (NSERC)**, for providing funds to continue lab research.

Contribution to Original Knowledge

Preface

This thesis is comprised of the work carried out during the doctoral studies and is presented in the form of manuscripts. In this study I have used advanced OMICS techniques and transcription factor analysis to map the transcription factors of high importance. To identify these genes, metabolites and transcription factors, we compared two potato genotypes Libertas and AG704 which is an advanced breeding line developed by Mr. Andre Gagnon. We further conducted gene characterization and validation to identify the reason for difference in resistance in the two genotypes.

Contribution to knowledge

- The transcriptomic study of two potato genotypes, AG704 and Libertas, revealed several resistance related genes upregulated in resistant genotype but were missing in the susceptible genotype post pathogen inoculation.
- The metabolomics of these genotypes also identified metabolites belonging to pathways which play a role in conferring resistance against late blight to potato.
- The transcription factor enrichment analysis comparing the two genotypes further identified the high impact transcription factors, which play a crucial role in transcriptional regulation post pathogen infection.
- A two-tier transcription factor regulation map was deciphered from the transcription factor enrichment analysis, which further allowed us to identify the secondary transcription factors.

- Transcription factors were also linked to the potential metabolites which they regulate through transcriptional regulation in the resistant genotype which are missing in the susceptible.
- Also, hydroxycinnamic acid amide (HCAA), feruloylagmatine, was identified which was highly accumulated in the resistant genotype and missing in the susceptible and was deciphered to be important for resistance.
- This was the first attempt to study the *ACT* gene in potato and several CDS and promoter sequences were obtained which are available to study and the *StACT2* was deciphered as the functional form of *ACT* gene in potato.
- The transcriptomic and metabolomic data is available for the research community to further study late blight resistance in potato which is an important disease.

Contribution of Authors

In this thesis two studies have been conducted (chapter III and chapter IV). This thesis is formatted according to the manuscript format suggested by McGill University, for the Doctoral thesis submission. All the greenhouse studies, pathogen inoculation, and metabolomics studies have been completed by me. Russiachand Singh Heikhm helped with the *de novo* transcriptomic data analysis and the further work was carried out by me. All the tissue culture, genome editing and *in silico* data analysis was carried out by me and Dr. Niranjan Hegde helped with the disease severity and tissue culture experiments.

The contribution of the authors is mentioned in each chapter mentioning the individual contributions in the connecting text.

Chapter I

Introduction

1.1 Introduction

Food security is the biggest task for the world in the next decades. It is a challenging road ahead, with the growing population and ever-increasing demand for food. Hence the major focus today is to increase yield and reduce the crop loss. Potato (*Solanum tuberosum* L.) is one of the major non-cereal staple food crops grown across the world. Recognizing the importance of potato as an important crop and a major dietary source to the world, United Nations Organization declared 2008 as the International Year of the Potato (http://www.potato2008.org). It is the highest grown staple crop after wheat, rice and maize (Birch et al., 2012). The total production of potato worldwide was 368 million tonnes in 2018 (FAOSTAT data, May 2020). Canada produces approximately 5.7 million tonnes of potato every year (FAOSTAT data, May 2020).

The major constraint in potato production is biotic and abiotic stresses. Among the biotic stresses late blight of potato caused by *Phytophthora infestans* continues to dominate (Fry et al., 2015). *P. infestans* is an oomycete, which can destroy the entire plant and can transform a healthy potato field into an entirely dead one, within a week (Fry, 2008). This pathogen possesses a great potential of causing an epidemic, the famous being the one in Europe in 1845 which lead to Irish famine (Fry, 2008). Recently Papua New Guinea faced an epidemic which lead to serious economic losses to farmers wiping out an entire crop (Price, 2004). Farmers in West Bengal state of India also faced similar losses due to introduction of a new clonal lineage of the pathogen which led to an epidemic (Chowdappa et al., 2013; Fry, 2016). Even North America faced severe late blight disease in 1994-1995 and again in 2009 (Fry and Goodwin, 1997; Fry et al., 2013).

P. infestans has a worldwide occurrence, leading to yield loss of up to 40% and hence is one of the most destructive plant diseases. It is estimated that the loss due to the disease can feed several million people (Fisher et al., 2012). Late blight is a serious problem in cold and humid conditions. If these conditions occur early in the cropping season the loss is up to 80%, in spite of applications of fungicides (Nowicki et al., 2012). Around the world the late blight is managed based on application of fungicides (Haverkort et al., 2009), but it can lead to development of resistant isolates (Gisi and Cohen, 1996; Gisi et al., 2011). Environmental impacts of fungicide are also a

reason to worry, which can have adverse effect on human habitation and health (Fisher et al., 2012). Forecasting systems are used to reduce the number of fungicide sprays based on weather conditions which are favourable or unfavourable for the late blight outbreaks (Fry, 1977; Hijmans et al., 2000; Filippov et al., 2015; Small et al., 2015) but, still the number of sprays required is high. Hence, integrated disease management is considered the best way to manage late blight (Mundt et al., 2002). Genetic resistance in the form of resistant varieties are considered to be the most economical and ecofriendly way to tackle such an epidemic pathogen (Haverkort et al., 2009).

The pathogen produces effectors to attack plants which are recognized by trans-membrane spanning, pathogen effector recognition receptors genes (Chisholm et al., 2006). This recognition is very specific and led to the gene-for-gene hypothesis of plant disease resistance (Flor, 1971). However, these are not really the resistance genes but are just surveillance genes that perceive the pathogen signal (Kushalappa et al., 2016a). Several receptor genes have been identified in potato and used in breeding programs to improve resistance to late blight (Song et al., 2003). The genetic flexibility of the pathogen, *P. infestans* has led to the failure in achieving this in most cases (Fry, 2008). The cultivated potato is an autotetraploid and because of sexual incompatibility the breeding has been slow or mainly focused on diploids (Slater et al., 2014). The cultivated potato genotypes have been crossed with several wild species, such as *Solanum demissum, S. bulbocastanum, S. stoloniferum, S. venturii, S. americanum* (Song et al., 2003; Jo et al., 2014; Witek et al., 2016). The progenies with reduced disease severity have been identified and developed as new cultivars, such as Defender and Premiere (Novy et al., 2006; Haverkort et al., 2009) although they are limited in diversity and yield. Though the genes involved, and mechanisms of resistance are unknown these have exhibited reduced infection efficiency, lesion size and reduced the rate of disease progress. Along with specific genes, molecular breeders have identified hundreds of quantitative trait loci (QTL), or segments of chromosomes, that are associated with high levels of resistance. These QTL are mapped on all the 12 chromosomes of potato in segregating populations (Simko, 2002) and almost 211 QTL have been identified (Danan et al., 2011).

Plants can resist the pathogen in either qualitative or quantitative way. Qualitative resistance is generally expressed as hypersensitive response. Quantitative resistance on the other hand is

reduced susceptibility and is controlled by polygenes (Poland et al., 2009). Recently, a novel concept of resistance has been proposed. The resistance is defined as reduced susceptibility, varying from hypersensitive response (qualitative) to high disease severity (quantitative). The resistance is due to hierarchies of genes, including regulatory genes that regulate downstream genes to biosynthesize resistance related (*RR*) metabolites and proteins that directly suppress the pathogen progress (Kushalappa et al., 2016a). These *RR* metabolites and proteins are the actual end products of the cascade which starts with the recognition of the pathogen by receptors that trigger the regulatory genes which regulate the genes that biosynthesize downstream *RR* metabolites and the genes that code for *RR* proteins (Kushalappa et al., 2016a). The cultivars may have one or more hierarchies of genes, but some links in the chain may be missing. Thus, missing the specific sets of end products such as *RR* proteins and *RR* metabolites that give direct resistance. By replacing these mutated or non-functional genes with functional genes it is possible to improve resistance in potato against late blight (Kushalappa et al., 2016b). Furthermore, these genes are expressed under certain conditions and they are controlled by several regulatory factors like transcription factors and hormones. Transcription factors play a major role in orchestrating these genes for a specific task only when the signal is perceived like pathogen invasion in case of biotic stress.

Hence, in the present study, our objectives were to discover new genes that are functional in a resistant potato genotype Libertas but are polymorphic, mutated or non-functional in a commercial susceptible genotype, AG704. Also, to identify and map the transcription factors that regulate these genes. Then, to use these genes that are functional in Libertas to replace the non-functional genes in the commercial genotype AG704 based on genome editing, using CRISPR-Cas9 system. We also tested for the change in resistance phenotype when the polymorphic region in the resistance genes are replaced with functional copy based on resistant genotype.

1.2 General hypothesis

Resistance is governed by several resistance genes which regulate downstream genes to biosynthesize proteins and metabolites. Potato genotypes with contrasting levels of resistance against late blight will vary in their metabolite, protein and gene expression profiles. The genes

that are functional in resistant genotypes but mutated or non-functional in susceptible genotypes are associated with biosynthesis of higher amounts of antimicrobial proteins and metabolites. Identification and replacement of these missing links in the chain will increase resistance in susceptible genotype.

1.3 Objectives

1. To identify differentially accumulated metabolites and expressed genes through metabolomics and transcriptomics of AG704 (late blight susceptible) and Libertas (late blight resistant) potato genotypes.
2. To select one important RR metabolite and characterize the genes responsible for its biosynthesis and to edit this gene candidate in the susceptible genotype AG704, to enhance resistance.

Chapter II

Review of Literature

2.1.1 Food security, the current concern

The world today is seriously concerned of being food insecure. The growing population always demands more food production. Food security is one among the 17 goals set by the United Nations Organisation (UNO) for sustainable development (http://www.un.org/sustainabledevelopment/hunger/). It is only possible to achieve food security when the multidisciplinary approach is developed and employed. There are various issues related to food security such as food safety, climate change, malnutrition which also needs to be tackled (Karunasagar and Karunasagar, 2016). One of the most vital areas we need to focus on today is to increase the food production and minimize the losses (MacLeod et al., 2016). Plant health is important along with the human health as plants keep humans healthy and provides necessary nutrition. Ignoring the plant health can lead to famines and adversely affect the sustainability goals.

2.1.2 Potato: A part of daily diet

Potato (*Solanum tuberosum*) belongs to an economically important family of plants, Solanaceae. The other crops in Solanaceae include tomato, tobacco, eggplant and pepper. It is believed to be originated in the Andes, in South America and is a tuber crop which feeds most people with dietary requirements. According to one study potato production is found responsible for the increase in population and urbanization for the past 2 centuries (Nunn and Qian, 2011). Hence it has always been one of the most important food resources since its domestication in South America.

Average consumption of potato worldwide is 33 kg/capita/year (FAOSTAT, May 2014). Potato is a versatile, carbohydrate-rich food highly popular worldwide and prepared and served in a variety of ways. Freshly harvested tubers of potato are made up of 77% water and the rest is dry matter which is predominantly starch. The potato tubers are very low in fat but rich in several micronutrients, especially vitamin C, riboflavin, thiamine and niacin; minerals such as phosphorous, potassium, magnesium and calcium. (Nutrition information from United States Department of Agriculture, National Nutrition Database). As potato is one of the important crops which feeds the world, it is necessary to reduce crop loss along with increasing yield.

2.1.3 Notorious oomycete: *Phythophthora infestans*

Phythophthora infestans is a pathogen well known for its destructive nature and known to infect potato and tomato plants (Agrios, 2005). It is regarded as one of the most destructive disease today and most devastating when the weather is cool and moist (Fry et al., 2015). The origin of *P. infestans* is under debate and 2 competing theories exist which place the origin in the Central Mexico or in the South American Andes (Goss et al., 2014; Saville et al., 2016). It is evident that the first reported outbreak of late blight was in 1843 on the east coast of United States of America particularly in Philadelphia and New York. The disease then crossed the Atlantic Ocean and was reported in Belgium where it spread through a shipment of infested seed potatoes (Nowicki et al., 2012). When the pathogen reached Ireland, which was a country mainly dependent on potato as a farm crop, it created an epidemic. The outbreak of *P. infestans* in Ireland led to a famine which lasted from 1845 to 1849. The "Great Famine" as it was called led to the death of one million people and forced another million to migrate to many places including the states in North America (Yoshida et al., 2013). The spread of the pathogen was never mitigated which resulted in a worldwide distribution of late blight in subsequent years (Fry et al., 2015). Even to this day, one hundred and fifty years after the great famine late blight leads to an annual loss of approximately $6 billion (Haverkort et al., 2009). The economic losses can be due to reduced yield, poor quality of tubers or even increased cost associated with fungicide applications.

P. infestans was originally termed as fungus, but later it was classified as an oomycete belonging to the kingdom Chromista. The oomycetes are evolutionarily placed closer with algae and other higher eukaryotic organisms than with true fungi. They have diploid genome in a well-formed nucleus and are coenocytic and lack chitin in the cell walls. The pathogen has a mycelial body type producing zoospores and sporangia as their asexual spores (Agrios, 2005). Sporangia are produced at the tips which are lemon shaped. The pathogen is a heterothallic with A1 and A2 mating types and produce oospores as their resting spores. In the past there was only one known mating type A1 in North America but, the other matting type A2 was identified in 1990's, increasing sexual reproduction eventually increasing the number of races (Peters et al., 1998; Fry et al., 2015).

The pathogen follows a hemi biotrophic mode of nutrition where initial infection is by a biotrophic mode and later the pathogen turns towards necrotrophic phase. In the early phase of infection (biotrophic) pathogen requires living host plant cell and derives the nutrition through the haustoria.

Zoospores are released from the sporangia which are typically three to eight in number at favourable temperature of up to 15°C. Sporangia may germinate directly by producing a germ tube when the temperatures are above 15^0C (Agrios, 2005).

The genome sequence of *P. infestans* is 240 megabases and encodes for approximately 18 thousand genes. The genome shows a unique character with sparse gene density. The repetitive genome almost covers one third of the genome. The effector genes are distributed in the repeat rich region whereas the housekeeping genes are mainly located in the repeat-poor region of the genome (Haas et al., 2009).

2.1.4 Integrated disease management practices

Since the first late blight outbreak, various types of management practices have been developed. Potato late blight management is a costly affair and most of the cost is incurred by the application of fungicide which are extensively used worldwide (Fry et al., 2015). The fungicides can be mainly of two types based on the way they work- protectant and systemic. Most commonly used fungicides are chlorothalonil, dithiocarbamates, triphenyl tin hydroxide, metalaxyl/mefenoxam, aliphatic nitrogen fungicides such as cymoxanil, morpholines such as dimethomorph (Nowicki et al., 2012). The application of fungicides is an effective measure until selection pressure is developed on the pathogen and pathogen overcomes the resistance (Peters et al., 1998; Gisi et al., 2011). Fungicide applications are found to be more effective when complemented with the weather forecast data (Fry, 1977). The pathogenicity of *P. infestans* is largely dependent on the moist weather with the relatively colder temperatures. Thus, algorithms have been developed to predict the conducive weather conditions to plan fungicide applications (Fry, 1977). It saves the excess fungicide application which otherwise can be costly and harmful (Hijmans et al., 2000; Filippov et al., 2015).

Various other measures are utilized to control or avoid the pathogen attack and eliminate the inoculum sources. The seed tubers should be disease free and they should be handled carefully to reduce or eliminate chances of infection by avoiding moisture conditions, avoiding plant debris or other potato field soils which is ideal for pathogen spore germination (Wale and Cattlin, 2008). The seed tubers can be treated with fungicides to eliminate spores, but it does not prove to be highly effective. Crop rotation and soil treatments by physical and/or chemical measures may

prove to be a good elimination strategy (Wale and Cattlin, 2008). Use of certified seed tubers may help to reduce the risk of disease. Moreover, the problem becomes acute when tomato and potato are grown together, or in close vicinity (Chen et al., 2008).

2.1.5 Potato breeding for late blight resistance

The breeding for resistance has been a continuous process since the great famine in Ireland. The pathogen has now spread to almost all the continents and hence late blight resistance breeding is an ongoing program across the world (Fry et al., 2015). The breeding programs in potato and tomato are concentrated on both race specific and race non-specific resistance that is, vertical and horizontal resistance respectively (Haverkort et al., 2009). Plants with vertical resistance has been an area of prime focus in developing late blight resistant potato. This type of resistance is effective in protecting the crop initially, but rapidly evolving pathogen effectors and sexually reproducing pathogen, overcomes with more aggressive lineages (Sujkowski et al., 1994; Fry, 2008). On the other hand, horizontal resistance or race non-specific resistance is durable over a long period and reliable which are controlled by polygenes or quantitative trait loci (QTL) (Thurston, 1971). The quantitative disease resistance controlled by polygenes is still not completely deciphered and hence less extensively used in breeding (St.Clair, 2010; Kushalappa et al., 2016a).

Various QTL which are effective in providing field resistance against late blight have been identified. These QTL are identified based on several cultivated and wild populations with interspecific and intraspecific crosses (Collins et al., 1999; Śliwka et al., 2007). Many of the wild populations of the genus *Solanum* have been crossed with *Solanum tuberosum* which contributed to several loci which have are effective against late blight. These include *Solanum demissum, S. bulbocastanum, S. papita, S. stoloniferum* and others (Haverkort et al., 2009). A cross between cultivated species *S. phureja* and *S. tuberosum* have also been effective in deciphering a few loci. A cross between a susceptible (*S. spegazzinii*) and resistant (*S. chacoense*) led to identification of new QTL on chromosome 4 and 5 for LB resistance (Chakrabarti et al., 2014).

Along with the QTL, individual race specific late blight resistance genes have been identified in potato wild species. There are 11 race specific genes named *R1* to *R11* in *S. demissum* alone (Bradshaw et al., 2006). Many of these major genes have been cloned and proven to be effective against foliar resistance. These genes belong to CC-NBS-LRR (coiled coils- nuclear binding site-

leucine rich repeats) protein family (Ballvora et al., 2002). Three other important genes are identified in *S. bulbocastanum* conferring broad spectrum resistance (Song et al., 2003). *RB* is a very important gene among these which was mapped to chromosome 8 and cloned (Song et al., 2003; Bhaskar et al., 2009). The problem is utilizing this gene in breeding programs, as wild species *S. bulbocastanum* cannot be directly crossed with the cultivated potato. Another homologue of the same gene has been identified in another species *S. stoloniferum*, which can be used in breeding (Wang et al., 2008). Though *RB* gives broad spectrum resistance, isolates which can be virulent and overcome *RB* resistance have been identified (Champouret et al., 2009). In 39 wild species of potato 17 new *RB* homologues have been identified which can be an important genomic resource for gene discovery (Tiwari et al., 2015). In a recent study 184 tetraploid cultivars were used to find 9000 single nucleotide polymorphisms (SNP's) and used to unravel several candidate genes like 3-hydroxy-3-methylglutaryl coenzyme-A reductase and P450 protein (Mosquera et al., 2016).

2.2.1 Metabolomics technology

Metabolites are accumulated in plants and perform various functions from structural components to defence. Taking into consideration the sessile nature of plants, the metabolites in various forms becomes paramount for growth, signalling, communication, defence, stress annihilation and overall resource allocation (Matyssek et al., 2002). Metabolomics has played a significant role in filling up the gap between phenotype and genotype and helping to make full genome sequence annotation a reality (Hall et al., 2002; Hall, 2006).

Metabolomics like other omics technologies, is a systematic identification, quantification and study of all the metabolites of an individual organism under specified conditions (Shulaev, 2006). Metabolic profiling can help answer several questions related to the plant phenotypes and it can form the missing link between the genotype and the phenotype, where metabolites being closer to phenotype (Fiehn, 2002; Hall et al., 2002; Okazaki and Saito, 2012). The metabolomics has advanced but considering its potential in unravelling biological questions it is still to advance to the level of other omics technologies like genomics and proteomics. The bottle neck being the difference in the fundamental chemical structure (Okazaki and Saito, 2012). Despite the limitations, metabolomics has been widely exploited as a tool for functional genomics, discovering biomarkers, safety assessment of genetically modified crops, QTL analysis, stress resistance

(abiotic and biotic) and trait improvement in plants (Soga et al., 2006; Urano et al., 2009; Ward et al., 2010; Gunnaiah et al., 2012; Yogendra et al., 2014; Yogendra et al., 2015b). This is an ongoing journey in metabolomics which has to still go a long way and we need more research in metabolomics to reach an exhaustive desired destination where every plant metabolite is identified (Hall, 2018).

2.2.2 Metabolite extraction and analytical platforms

The metabolome consists of a vast array of compounds; hence the fully integrated strategy for metabolite extraction, optimal separation, detection, identification, data analysis will affect the final output and advancement in these steps in turn dependent on analytical and computational developments (Hall et al., 2002). Metabolite extraction is the first important step in metabolic profiling studies and must be comprehensive. It should also possess an ability to quench maximum metabolites. Apparently, no single solvent can extract all the metabolites and hence one which gives close to maximum extraction should be selected (Shulaev et al., 2008; Kushalappa and Gunnaiah, 2013). The selection of metabolites also depends on the polarity of the analytes in consideration and the analytical platform. Several concentrations of methanol and chloroform have been tried and tested which shows 70% to 75% methanol (v/v) can give the higher number of metabolites. It was found most efficient in extracting wide range of secondary metabolites from various plants and tissues (De Vos et al., 2007).

In the past two decades, various platforms for large scale metabolomics have been introduced, which have the potential to analyse all the metabolites. Every platform can perform certain things better than the other and plant metabolites due to their wide range in concentration and molecular weight make it complicated (De Vos et al., 2007; Okazaki and Saito, 2012). There is a need to increase the metabolome information and make it more comprehensive by integrating various analytical platforms depending on to the chemical properties of the metabolites (Shulaev, 2006). For example, mass spectrometry (MS) is often used in metabolic profiling studies coupled with separation technique like chromatography. This combination increases the number of compounds detected by minimizing the complexity of mass spectra (De Vos et al., 2007; Okazaki and Saito, 2012). Several chromatography platforms can be utilized based on cost effectiveness, sensitivity and best suited for specific purpose. Liquid chromatography (LC) and gas chromatography (GC) are widely used separation techniques in metabolite studies having some pros and cons over each

other. Other platforms include direct flow injection (DFI-MS), capillary electrophoresis (CE-MS) and nuclear magnetic resonance (NMR) techniques (De Vos et al., 2007).

GC coupled with MS based metabolite profiling serves as a highly sensitive reliable, robust but its limited to volatile compounds and requires derivatization (Lisec et al., 2006). It also has an advantage of large commercial and public libraries being available (Schauer et al., 2005). GC in combination with time of flight (TOF)-MS was the popular technique and later GC-MS was routinely used (Fiehn et al., 2000; Lisec et al., 2006). CE-MS also serves the similar purpose like GC-MS but is not very popular and seldom used. It provides large scale quantitative data for tonnes of metabolites similar to GC-TOF or MS (De Vos et al., 2007).

LC-MS on the other hand is the most preferred technique for analysis of plant secondary metabolites. The advantage is that no derivatization is required, and it holds large sample capacity but, it is slow compared to other techniques (De Vos et al., 2007). LC in combination with a soft ionization technique can be more efficient. Electrospray ionization (ESI) or atmospheric pressure chemical ionization (APCI) is used which results in a positive or negative mode ionization. The limitations in terms of reference libraries are existing, but past few years' libraries are seen increasing exponentially (Tohge and Fernie, 2010; Kushalappa and Gunnaiah, 2013). LC-MS approach helps to identify and quantify many semi-polar compounds, which include the key secondary metabolite groups such as phenolic acids, phenylpropanoids, flavonoids, alkaloids, glycosylates, saponins and others (De Vos et al., 2007). Also, several polar metabolites depending on the column used for LC. Many important resistance related metabolites upon *Fusarium* infection in barley and wheat using LC-MS has been reported (Bollina et al., 2010; Bollina et al., 2011; Kumaraswamy et al., 2011a).

NMR is another platform which is non-destructive, non-biased, highly quantitative, does not require derivatization or separation, can determine the atomic state of compounds and enables identification of complex unidentified compounds (Okazaki and Saito, 2012). The drawback still lies in low sensitivity which reduces the utilization of NMR over mass spectrometry (De Vos et al., 2007). Some other direct injection approaches like Fourier transform-ion cyclotron resonance mass spectrometry (FTMS) and TOF-MS have been used for metabolite fingerprinting without separation. FTMS has been a platform of choice due to high sensitivity and resolution. FTMS holds two major advantages over other analytical platforms; ability to detect compounds prior to

separation and provides accurate chemical formulae of the detected peaks which helps in metabolite annotations (Okazaki and Saito, 2012). However, direct delivery systems may increase the adduct formation and it lacks detection system for molecular isomers (De Vos et al., 2007). Matrix-assisted laser desorption/ionization (MALDI) imaging has also been used to detect plant metabolites but with limited applications (Kaspar et al., 2011). Also, platforms like stable isotope-enabled mass spectrometry has opened new doors for plant metabolomics which are exploited to a greater extent and have been advantageous in improving the metabolite annotation and pathway identification (Freund and Hegeman, 2017).

2.2.3 Computational framework and advancement of metabolomics

Data processing is the key step in large scale, untargeted metabolomics. Several tools are available for processing LC-MS data and depending on the use and scale the right program can be chosen. The drawback remains in the metabolite identification, as very few secondary metabolites in plant species have been identified compared to the primary metabolites (Kushalappa and Gunnaiah, 2013). The development in the past decade has made the data processing holistic, unbiased and made it simple than before. Several software packages are available for mass peak detection and alignment like XCMS (various (X) forms of chromatography MS) (Colin A. Smith et al., 2006), $XCMS^2$ (Benton et al., 2008), MetAlign (Lommen and Kools, 2012), MZmine (Katajamaa et al., 2006; Pluskal et al., 2010), Markerlynx and others. These packages involve four basic steps: deconvolution, grouping, alignment across samples and gap filling. Before alignment, analytical information in the profile is transformed into coordinates based on mass identified and retention time (De Vos et al., 2007). The output of these software packages often contains multiple peaks for same compound along with adducts, isotopes, and dimers. The metabolite annotation/identification is done using the accurate mass, fragmentation pattern and number of carbons based on isotope ratio (Kushalappa and Gunnaiah, 2013). The accurate mass of compounds can be referenced with the masses of compounds in public or commercially available libraries and databases. The fragmentation patterns can be matched like accurate mass, with the databases like KEGG, METLIN, Lipid Maps, PlantCyc, MetaCyc, MASS BANK, KNAsSAcK and others. The fragmentation patterns can also be verified manually using ChemDraw, ChemSketch or other chemical drawing software (De Vos et al., 2007; Kushalappa and Gunnaiah, 2013). The fragmentation patterns can be also confirmed if standards for the compounds are

available. The number of carbon atoms present in the molecular based on isotope ratio can be used to identify the metabolites (Kushalappa and Gunnaiah, 2013).

2.3 Transcriptomics

Since the sequencing of genome is a reality, and high throughput sequencing techniques have come up, new ways to understand the biological systems are found. The total quantity of RNA molecules present in a biologically active sample is referred to as transcriptome and the study of transcriptome is known as transcriptomics. The transcriptome is highly active and is dynamic to all the stimuli and cellular activity (Wang et al., 2009). To study the transcriptome there was high dependency on DNA microarray technologies. More advancement followed by techniques like serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE) and other massively parallel sequencing (Wang et al., 2009). RNAseq has gained enormous popularity due to numerous advantages over other approaches. It is not completely dependent on the existing genomic sequences and novel gene/transcript identification is possible. There is a possibility of large dynamic range of detection very low fold difference is quantifiable and is precise (Wang et al., 2009).

Transcriptome sequencing produces huge amount of data and analysis of it is challenging like other high throughput techniques. The analysis can be done in two ways; either by aligning the reads to a reference genome or by *de novo* assembly of the reads to produce genome-scale transcription map (Trapnell et al., 2012). The reference-based analysis is divided into three major steps. First reads are mapped to the available reference genome or transcriptome which is followed by assembling the transcript. The assembled transcripts are then quantified for expression level. Several alignment programs are freely available which can map the reads to the reference, such as bowtie (http://bowtie-bio.sourceforge.net/index.shtml) or TopHat (http://ccb.jhu.edu/software/tophat/index.shtml) are popular and widely used. Taking into consideration the limitation of bowtie, that it cannot align reads with introns, TopHat will be the best choice for alignment. These programs can only be used for sequencing platform reads from Illumina and SOLiD, and TopHat can additionally detect transcript splice sites (Trapnell et al., 2012). Once the mapping of reads is done it must be assembled into transcripts. Cufflinks (http://cole-trapnell-lab.github.io/cufflinks/) is a widely used computational tools for transcript assembly. With Cufflinks, the mapped reads from TopHat are first assembled to estimate their

abundances. The Cuffmerge assembles several RNAseq libraries into a master transctiptome. Cuffdiff which is part of the Cufflinks package tests for differential expression and regulation in RNAseq samples (Trapnell et al., 2012). The output from Cuffdiff can be visualized using CummRbund. Along with Cufflinks package other programs like DegSeq, EdgeR, differential expression analysis of count data (DESeq) are also commonly used for identifying differential gene/transcript expressions (Oshlack et al., 2010).

A monoploid potato genome was recently sequenced and the assembly counts to about 726 MB, out of which 86%, is aligned to genetic map. This constitutes around 39 thousand genes (Xu et al., 2011; Hirsch et al., 2014). The reference genome sequence of potato (The Potato Genome Sequencing Consortium- http://www.potatogenome.net/index.php/Main_Page) was obtained from a double monoploid (DM-1) and additionally as a second line from heterozygous diploid line (S. tuberosum Group Tuberosum RH89-039-16) (Xu et al., 2011). Along with the whole genome assembly RNAseq data from several tissue types and developmental stages are available in the sequence consortium (http://solanaceae.plantbiology.msu.edu/pgsc_download.shtml). These resources will be useful to analyze the potato RNAseq data for assembling reads.

2.4.1 Plant resistance mechanisms

Plants employ several strategies to contain the invading pathogen. It involves both resistance related (*RR*) proteins and metabolites. The resistance imparted can be constitutive or induced depending on the protein expression and metabolite accumulation in plants. These *RR* proteins and *RR* metabolites either suppress the pathogen or deposited as structures to reinforce the cell walls and contain the pathogen to infected area (van Loon et al., 2006).

Several pathogenesis related (PR) proteins have been identified in different plants which can directly or indirectly supress the invading pathogen and they are grouped in 17 different families (Van Loon and Strien, 1999; van Loon et al., 2006). The PR proteins possesses the antimicrobial, cell wall penetrating and toxin- degrading properties (Nawrot et al., 2014). Antimicrobial proteins (AMP) exert antimicrobial activity against different pathogens by interacting with the phospholipids and metabolic processes which makes the membrane permeable, ultimately killing the pathogen (Eudes and Chugh, 2008). The prominent families of PR proteins comprise defensins, thionins, lipid transfer proteins, snakins, cyclotides, knottins and hevein-like peptides (Nawrot et

al., 2014). The other families like PR-1 and PR-5 (osmotins) are mainly important against the oomycetes, which act as membrane permeabilizing agent (van Loon et al., 2006). The other important class of PR proteins called extensins play an indirect role in containing pathogen by thickening the call wall and making a barrier for pathogen progress (Łaźniewska et al., 2012). Some proteins may also help in reducing the toxic effects of the invading pathogen, like in the interaction between *Fusarium graminearum* and barley. The toxic compound deoxynivalenol produced by the pathogen is glycosylated to reduce the toxicity by DON-3-glycosyl transferase enzyme (Schweiger et al., 2010).

Along with the proteins, secondary metabolites play a vital role in defending the invading pathogen. The biosynthesis of specific metabolites is one of the most powerful defence reactions plants employ (Piasecka et al., 2015). Plants in natural habitats are attacked by several pests and pathogens which impair growth and development and lead to loss in reproductive capabilities (Dangl and Jones, 2001). The metabolites in the structural barriers like trichomes and cuticle fortify the invading pathogen (Łaźniewska et al., 2012) which are collectively called as phytoanticipins. The phytoanticipins are constitutive and nontoxic, but toxic forms are released upon hydrolysis, like the saponins, glucosinolates, cyanogenic glucosides and benzoxazinone glucosides (Piasecka et al., 2015). Plants produce several other groups of metabolites including phenols, fatty acids, terpenoids, flavonoids, alkaloids which can also be induced upon pathogen attack and called phytoalexins (Ahuja et al., 2012).

2.4.2 Metabolomics for disease resistance in plants

The elucidation of defence mechanisms with a metabolomics approach can help in understanding the strategies of plants against stresses (Okazaki and Saito, 2012). Several metabolites have been identified for disease resistance in different plants. Metabolomics of plants with contrasting levels of resistance to pathogen stress have revealed several metabolites in wheat against fusarium head blight (Gunnaiah et al., 2012) and in potato against late blight (Pushpa et al., 2013; Yogendra et al., 2014; Yogendra et al., 2015b). Resistance related metabolites have been identified in wheat and barley upon *Fusarium graminearum* infection which can be used as biomarkers in high throughput screening of breeding lines (Hamzehzarghani et al., 2005; Kumaraswamy et al., 2011b). Resistance due to the deposition of hydroxycinnamic acid amides in the cell wall, preventing the pathogens to penetrate and colonize have been identified in wheat and potato

(Gunnaiah et al., 2012; Pushpa et al., 2013; Yogendra et al., 2014; Yogendra et al., 2015b). Untargeted metabolomics approach has been used to identify *RR* metabolites and genes in potato and reported the deposition of hydroxycinnamic acid amides (HCAA's) at host cell walls inhibit pathogen colonization imparting resistance against *P. infestans* (Pushpa et al., 2013). Also, the genes involved in regulation of these HCAA's are reported in potato against *P. infestans* (Yogendra et al., 2014; Yogendra et al., 2015a). In a recent study, higher accumulation of morphinone, codeine-6-glucuronide and morphine-3-glucuronides was noticed in resistant potato compared to susceptible upon *P. infestans* pathogen inoculation (Yogendra et al., 2017a).

2.5 Transcriptional reprogramming during resistance response

Plant disease response is a synergistic response where several metabolites and genes play a significant role. To better understand the disease mechanism in plants it is very important to understand how it is controlled and regulated. Several transcription factors play a major role in controlling the disease response in plants (Meraj et al., 2020). Along with transcription factors hormones and small RNAs also play a role in orchestrating the response in couple with the transcription factors (Zhang et al., 2013; Wiesel et al., 2015). Mapping of important disease resistance metabolites on to metabolic pathway have identified their biosynthetic genes and silencing of these genes have confirmed the resistance roles of these genes (Yogendra et al., 2015a). Metabolomics and transcriptomics studies have identified several hierarchies of genes against late blight (Yogendra and Kushalappa, 2016). Various transcription factors (TFs) are found to be controlling the downstream genes responsible for quantitative resistance. *WRKY* is one the important TF found in potato, which regulate the downstream phenylpropanoid pathway (Yogendra et al., 2015a). The important transcription factors like *MYB*, *NAC*, and *WRKY* are actively participating in upregulation of genes that biosynthesize *RR* metabolites (Kushalappa et al., 2016a). However, most of these gene functions have been only characterised in model plants and the discovery of their homologues genes in potato source genotypes can enhance their use in breeding programs.

2.6.1 Genome editing using CRISPR Cas9

Genetic manipulations in last century, which focused on uncertain chemical and physical mutagens has now shifted to a new paradigm in which we can pinpoint specific locations to edit (Huang et

al., 2016). As the number of genome sequences available is increasing, there is a growing understanding that the genome of a single individual differs in gene diversity compared to the other individuals in species. This leads to the phenotypic variation amongst members of a species and based on these phenotypes they are been selected. Genome editing in the form of insertions, deletions or replacement can thus help to achieve a desired phenotype keeping the other functional genotype as reference.

The recent development in precise nucleases targeting specific loci in the genome of interest in the form of Meganucleases (MN's), Transcription activator-like effector nuclease (TALENS), Zinc-finger nucleases (ZFN's) and Clustered regularly interspaced short palindromic repeats/CRISPR-associated protein (CRISPR/Cas) have given hopes to achieve desired manipulations. CRISPR/Cas9 genome editing system has revolutionised the editing with precise, easily programmable option and said to be the next big thing in the world of science. Very soon the CRISPR turned into a craze and it was taken to by every possible application from bacteria to humans and plants (Pennisi, 2013). Its use in editing plant genomes has been extensively tried and tested to achieve desired traits since its inception. The technology which was first identified in bacteria and archaea as an immune system was developed as a tool for programmable editing (Jinek et al., 2012). In a CRISPR-Cas system, both crRNA and trcrRNA components are linked together through what is called as a 'linker loop,' thereby making the system easy to manipulate and use (Jinek et al., 2012).

CRISPR/Cas provides an option to have a double strand break (DSB) at a desired site based on the sequence complementary to the CRISPR RNA (crRNA). These DSB's can be utilised to achieve gene or chromosomal segment deletions, inserting a segment of DNA or introducing specific nucleotide change by using the plants own DNA repair mechanisms (Voytas, 2013). The non-homologous end joining (NHEJ) or homology directed repair (HDR) in plant can make up the DSB to desired outcome (Baur et al., 1990; Puchta et al., 1996; Gorbunova and Levy, 1997; Voytas, 2013).

2.6.2 CRISPR-Cas9 and plant improvement

With the advent of genome editing technology its applications in every area was explored. Plant improvement has always been a desire to achieve more yield, minimize losses and increase the

production. Hence CRISPR technology can be effective in obtaining plants with desired traits. Insertions of foreign gene at a desired location can help improve crops with desired traits, or at times even silencing a gene can be effective (Voytas, 2013).

The first report of CRISPR genome editing in plants started with the model plant, *Arabidopsis thaliana* (Feng et al., 2013) followed by *Nicotiana bhentamiana* (Nekrasov et al., 2013). Soon the technology was tested for many other commercially important crops and non-model plants including rice and wheat (Li et al., 2013; Shan et al., 2014). Various transformation techniques have been tried for efficient plant transformation for CRISPR, from *Agrobacterium* mediated, gene gun, to polyethylene glycol mediated transformation. Initially only plasmid-based editing was shown to be possible but soon the pre-assembled protein and crRNA or guide RNA (gRNA) was used with different transformation techniques (Woo et al., 2015; Liang et al., 2017).

With the development in genome editing technology new ways to make it more efficient and precise are always carried out. It is said to be very helpful in plant breeding and improvement, regardless of the regulatory concerns over the genetically modified crops and foods. Efforts are underway to keep genome editing using the latest technology, out of the conventional genetically modified category (Huang et al., 2016). The breeding programs have been mainly focused on yield improvement and hence, the cultivated crops render narrow genetic diversity. High yielding cultivars are further neglected when the disease prone behaviour is evident. Thus, by identifying the specific non-functional genes and making it functional can help improve a cultivar for disease resistance (Kushalappa et al., 2016b).

2.6.3 Advancement in CRISPR technology

Since the use of CRISPR Cas9 started it has advanced many folds to different arenas. Several other technologies previously known were combined with the CRISPR to further advance and harness maximum potential of these technologies. For example, plant viral replicons play a crucial role in gene silencing however, when they are combined with CRISPR they can add value to the outcome to enhance the homology directed repair (Anandalakshmi et al., 1998; Gil-Humanes et al., 2017). Also, constructing CRISPR vectors for plants were made easy using modern cloning techniques like the Golden Gate and Gateway® technologies. Several customisable toolkits were developed

which could be used for multiple plants based on needs and also the other CRISPR components that is gRNA and donor DNA were made customisable (Čermák et al., 2017; Liang et al., 2017).

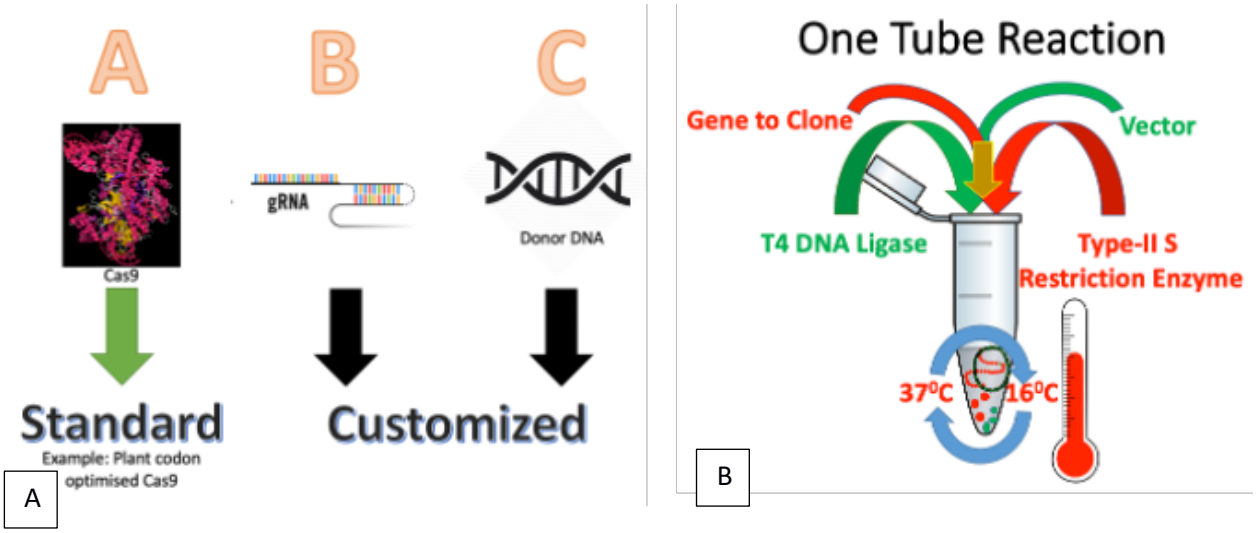


Fig 2.1 Representative image of a CRISPR toolkit layout

Shows the overview of the CRISPR Cas9 system with A, B and C components (A) and one tube Golden Gate cloning reaction (B) which helps in toolkit development. (Engler and Marillonnet, 2014; Čermák et al., 2017).

As off-targets in CRISPR has been a concern for plant community and it can have unknown effects in developing plants with undesired traits, identifying gRNAs with least off- target efficiency is crucial (Klein et al., 2018). Hence, the programs like CHOP-CHOP and CRISPR-P 2.0, which have the off-target identifier across several sequenced plant genomes have further enhanced the CRISPR technology (Labun et al., 2016; Liu et al., 2017). Also, other CRISPR associated nucleases like Cpf have helped in increasing the gene targeting efficiency many fold (Hu et al., 2017).

Connecting statement to chapter III

The chapter III is a manuscript titled **"Transcription regulatory maps reveal regulation of induced late blight resistance genes by transcription factors",** authored by Sripad Joshi, Russiachand Singh Heikham and Ajjamada C. Kushalappa.

Contributions of authors

Sripad Joshi conducted the experiments and wrote the manuscript. Russiachand Singh Heikham helped with the RNAseq analysis. Ajjamada C. Kushalappa supervised the project and edited the manuscript.

Potato late blight resistance is an ongoing struggle and several ways have been approached to deal with this problem. However, the understanding of the pathogen and plant interaction is incomplete and needs more thorough review. Several previous studies in potato late blight resistance have identified transcription factors which play a role in triggering certain metabolites which play a role in resistance. These include the *StWRKY8, StWRKY1, NAC43* and *MYB8,* however these studies have helped us understand certain pathways and genes (Yogendra et al., 2015a; Yogendra et al., 2017b; Yogendra et al., 2017c). However, in this study we have attempted to understand and map the broad-spectrum transcription factors which may have large scale impact over several genes. To do so, we used transcriptomics and metabolomics technologies to identify the upregulated genes and accumulated metabolites in a resistant and a susceptible genotype. The integration of metabolomics and transcriptomics have been attempted before, but in this study along with the integration we further carried out a transcription factor enrichment analysis, comparing the genes that are only upregulated in resistant genotype and missing in susceptible and further mapping them to the metabolites which are accumulated.

Chapter III

Transcription regulatory maps reveal regulation of induced late blight resistance genes by transcription factors

Sripad Joshi[1], Russiachand Singh Heikham[1], Ajjamada C. Kushalappa[1*]

Department of Plant Science, McGill University, Ste.-Anne-de-Bellevue, Quebec, Canada

3.1 Abstract

Late blight of potato (*Phytophthora infestans*) causes up to 40% yield loss worldwide. Resistance to late blight is either qualitative or quantitative. Even though quantitative resistance is durable, the mechanisms of resistance is not well deciphered, which limits its applications. In this study, several induced resistance genes were identified, which biosynthesize resistance related (*RR*) induced metabolites upon pathogen attack, in higher fold change in a resistant genotype (Libertas) than in a susceptible genotype (AG704). The pathogen induced gene expressions and metabolites biosynthesized were obtained by RNAseq and metabolomics, respectively. A total of 281 constitutive and 160 induced metabolites, belonging to different chemical groups were identified. RNAseq *de novo* assembly revealed 611 induced genes which were further categorized based on their biological functions. Promoter sequences of these genes were identified, which was used for regulation prediction and Transcription Factor (TF) enrichment study to identify regulatory TFs. A total of 134 TFs were found highly enriched in TF enrichment study and had binding sites in the promoters of several induced genes. A correlation study among the induced genes further mapped several primary and secondary TFs. Among these *bHLH66, MYB61, NAC56, WRKY51*, *MYB like, ERF RAP2-3* and *MADS-box AGL15* are predicted to regulate more than 208 downstream genes, of which many were secondary TFs. Hence, a two-tier transcriptional regulation of defense response genes was mapped. This study offers a deeper insight into the complex regulatory role of genes and their biosynthetic metabolites in disease resistance. Following validation, these genes can be used to develop disease resistant genotypes.

3.2 Introduction

Potato (*Solanum tuberosum*) is a major crop grown across the world. It is a non-cereal staple food crop and is the highest grown staple food crop after wheat, rice and maize (Birch et al., 2012). The potato production faces challenges in terms of biotic and abiotic stresses. Among the biotic stresses, dominates the late blight of potato caused by an oomycete *Phytophthora infestans* (Fry et al., 2015). *P. infestans* can destroy an entire plant and has the potential to transform a healthy potato field into an entirely dead one, within a week (Fry, 2008). This pathogen possesses a great potential to cause epidemics, the famous one being the Irish famine in 1845 (Fry 2008). *P. infestans* has a worldwide occurrence, leading to yield loss of up to 40% and hence is one of the most destructive plant diseases. To tackle such an epidemic pathogen, use of resistant varieties are considered to be the most economical and ecofriendly way (Haverkort et al., 2009). The genetic flexibility of the pathogen also makes the breeding resistant cultivars very challenging (Fry 2008).

Resistance in plants against pathogen attack can be categorized into qualitative and quantitative resistance. Qualitative resistance is generally expressed as hypersensitive response. Quantitative resistance on the other hand is reduced susceptibility and is controlled by multiple genes (Poland et al., 2009). The resistance is due to resistance related metabolites and proteins that directly suppress the pathogen progress (Kushalappa et al., 2016). These metabolites and proteins are the end products of the cascade (which includes MAP Kinase, transcription factors and hormones). The invading pathogens produces elicitors and effectors, which are recognized by the host and triggers other signalling, regulatory and biosynthetic genes to produce metabolites and proteins (Kushalappa et al. 2016).

The overall resistance response can be divided into 3 main divisions (Li et al., 2016). First line of defense includes the receptor genes which are broadly categorized into Effector triggered immunity (ETI) and PAMP (pathogen associated molecular pattern) triggered immunity (PTI) (Jones and Dangl, 2006). The pathogen signal perceived further transduced to the transcription factors (TFs) by various signal transducers like the mitogen activated protein kinases (MAPKs) and Ca^{2+} dependent protein kinases (CDPKs). These TFs form the second division where the transcriptional reprogramming happens which sends further signals to hormone regulation which act as the secondary regulators. The third division is the defense response genes which are further

activated by the TFs and regulated by hormones which form the actual defense attack to contain the pathogen (Dangl and Jones, 2001; Jones and Dangl, 2006; Kushalappa et al., 2016).

The pathogen induced metabolites are not only antimicrobial, but they are also deposited to reinforce the secondary cell walls to contain the pathogen to initial infection area (Yogendra et al., 2015a). The genotypes varying in their degrees of resistance may have differences in their transcriptional reprograming after pathogen attack. Thus, missing the specific set of end products, such as proteins and metabolites, which directly suppress the pathogen. Identifying some of these missing links in the chain of signal transduction events will help us to manipulate the specific targets which have a larger impact on improving resistance. It is also necessary to note in terms of resistance that genotypic background of the plant plays a crucial role in conferring long sustaining resistance (Shandil et al., 2017). Hence, it is very important to understand the underlying mechanisms which govern the resistance and not only the *R* gene mediated pathogen perception. A comprehensive study based on RNA seq and semi-quantitative metabolomics of moderately resistant and moderately susceptible genotypes have been used to explore the regulatory roles of these TFs to trigger gene expression and biosynthesis of induced metabolites, to contain the invading pathogen.

In this study an attempt was made to underpin the regulatory framework which is secondary to the pathogen signal perceived by plant receptor genes and transduced to through the MAPK, CDPK and Ca^{2+} signalling (Li et al., 2016). Underneath the signalling mechanism lies the transcriptional reprogramming network which is the most important segment of the disease resistance regulation (Tsuda and Somssich, 2015). Further to explore the mechanisms regulating the quantitative resistance, the co-expressed TFs were linked with the genes which bring about the metabolic changes required for quantitative resistance. These TFs play a major role in the transcriptional reprogramming and metabolic reflux, post pathogen attack leading to a synergistic response via multiple pathways.

3.3 Materials and methods

3.3.1 Plant and pathogen production, inoculation and disease severity

Resistant genotype Libertas (obtained from Potato Gene Resources, Fredericton Research and Development Centre, Agriculture and Agri-Food Canada (AAFC), New Brunswick, Canada) and the susceptible genotype AG704 (obtained from Mr. Andre Gagnon, Progest 2001 Inc.) were used in this study. Either complete tubers or tubers with at least two eyes sprouted were planted in a pot with promix BX®, soil and perlite. These pots were maintained in a greenhouse at temperature 23±3^{0}C with 16 hours' light, 8 hours' dark photoperiod and 70±10% relative humidity.

P. infestans isolate US-8, A2 mating type (obtained from Dr. Rick Peters, AAFC, Charlottetown, Canada) was maintained on potato dextrose agar (PDA) media. The spores were produced by inoculating thin potato tuber slices and incubating in petri dish with enough moisture for growth. The culture was maintained at 18^{0}C until sporulation. Freshly erupted sporangia were harvested and suspended in water and the spore concentration was adjusted to 1x10^5 sporangia ml^{-1}.

The experiments were conducted in the greenhouse as randomized complete block design (RCBD), with two genotypes: AG704 and Libertas, and two inoculations: mock (sterile water) and pathogen (*P. infestans* spore suspension), with five replicates over time. Each experimental unit consisted of five pots with two plants in each pot. Young leaves and stems of 5-week-old plants were then point inoculated with 10 µl of suspension on either side of leaf midribs and on opposite sides on the surface of stems. Ten leaves and five stems were inoculated per pot. The pots were covered with plastic bags, sprayed inside with water, for 3 days. The leaflets and stem cuttings were harvested at 2, 3 and 6 dpi (days post inoculation), the disks containing lesions were cut with cork-borer, the samples were flash frozen in liquid nitrogen and stored at -80^{0}C. The disease severity was assessed by measuring the lesion diameter with a digital caliper, at every three day-intervals (3, 6, 9 dpi) until 9 dpi. From the lesion diameter, the lesion area and the area under disease the progress curve (AUDPC) was calculated (Fry, 1978; Jeger and Viljanen-Rollinson, 2001).

3.3.2 Metabolite extraction and LC-high resolution MS/MS analysis

Leaf samples collected at 3 dpi were weighed and ground with the pre-chilled mortar and pestle with liquid nitrogen. Metabolites were extracted from 5 replicates for each treatment using absolute methanol + 0.1% formic acid, followed by 60% methanol + 0.1% formic acid, to extract most of polar, semi polar and non-polar metabolites from the sample. Sodium taurocholate was used as an internal standard and the formic acid was used to increase the efficiency of ionization. These were analysed in a negative ionisation mode using a high-resolution, accurate-mass (HRAM) Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (LC-MS/MS) (Thermo Fisher, USA) using a 5 cm XB-C18 kinetex column. The samples were analysed in a randomised mode to avoid any structural errors associated with the LC-MS. The data files obtained from LC-MS/MS were converted to mzxml/.cdf format and data analysis was done using MZmine-2 (Pluskal et al., 2010) with peak deconvolution, peak detection, spectral filtering and normalization of peaks (Katajamaa et al., 2005).

The masses and the relative abundances were imported to spreadsheet and the monoisotopic mass (MS1) was separated from other multiple peaks, such as isotopes, adducts and dimers. A statistical comparison was made between two treatments at a time using monoisotopic peak abundance (ion counts) data which was subjected to students *t*-test to identify treatment significant metabolites.

Based on the higher abundance of metabolites in resistant genotype as compared to susceptible genotype, the resistance related (*RR*) metabolites were identified. The resistance related constitutive (RRC) metabolites were based on mock inoculations in resistant and susceptible (RRC= RM/SM>1.0) (Bollina et al., 2010). The pathogenesis related metabolite which has higher abundance in resistant as compared to susceptible were considered as resistance related induced (RRI) metabolites [RRI = (RP/RM>1.0) > (SP/SM>1.0)] (Gunnaiah et al., 2012). Where RM= resistant mock inoculated, RP= resistant pathogen inoculated, SM= susceptible mock inoculated and SP= susceptible pathogen inoculated.

3.3.3 RNA extraction and library preparation

Total RNA was isolated from 3 replicates using 100mg of leaf samples collected at 2 dpi, using the RNeasy Plant mini kit (Qiagen, Canada). RNA was quantified using NanoDrop

Spectrophotometer ND-1000. The integrity of the total RNA was assessed using the 2100 Bioanalyzer. Libraries were prepared from pooled samples of 3 biological replicates from each treatment using TrueSeq mRNA sample preparation kit (http://www.illumina.com/products/truseq_stranded_mrna_sample_prep_kit.html) as per the manufacturer's instructions. The mRNA was segregated using the poly-A selection and it was subjected to fragmentation. cDNA synthesis was done followed by ligation of adapters and cDNA fragments enrichment (PCR) (http://www.illumina.com/applications/sequencing/rna/mrna-seq.html). Further the libraries were quantified using Quanti-iT™ PicoGreen® dsDNA assay kit (Life Technologies, Canada) and the Kapa Illumina GA with revised primers SYBR fast universal kit (Agilent Technologies, Canada). Libraries were subjected to sequencing using Illumina Hiseq 4000 sequencer, with 100 bp paired end reads at the Genome Quebec, Montreal, Canada.

3.3.4 Transcriptomic data analysis

Sequencing data analysis was carried out using the *de novo* assembly. The high-quality output reads were preprocessed using ngsShoRT 2.2 (https://research.bioinformatics.udel.edu/genomics/ngsShoRT/index.html). The reads were assembled using 3 different programs *viz,* Trinity (kmer 25), Idba_tran (kmers 31 to 91 with 10 steps), SOAPdenovo-Trans (kmers 31 to 95 with 8 steps) (Grabherr et al., 2011; Peng et al., 2013; Xie et al., 2014). The cds based clustering was done using the EvidentalGene (http://arthropods.eugenes.org/EvidentialGene/) program. The differential gene expression was quantified using the accurate RSEM (Li and Dewey, 2011). The assembled transcripts were further annotated using the Annocripts (Musacchia et al., 2015). The assembled transcripts were further annotated with the potato reference genome to find the PGSC identifications (Xu et al., 2011).

3.3.5 Candidate gene validation using qRT-PCR

The total RNA was extracted from inoculated leaves 2 dpi, which was used for cDNA synthesis using the reverse transcriptase. The first strand was synthesized using AffinityScript QPCR cDNA Synthesis Kit (Agilent Technologies) and quantitative real time PCR was performed using IQ SYBR Green supermix (BioRad, Canada). The reaction was carried out in CFX384TM Real-Time

system (BioRad, Canada). The relative gene expression levels were analyzed using ΔΔCT (cycle threshold) method ($2^{-\Delta\Delta CT}$) along with the standard curve (Livak and Schmittgen, 2001).

3.3.6 Transcription factor enrichment analysis and regulation prediction

To understand the regulatory framework of the pathogen induced genes in the resistant genotype, TFs which have an impact beyond single gene were identified. To identify the regulating TFs, first the promoters of all the induced genes identified were extracted, using an in-house python script. The NCBI annotation (Sol_Tub_3.0) (https://www.ncbi.nlm.nih.gov/genome/?term=potato) and the transcript sequences available at NCBI were used. The list of the all the promoter sequences was then used to identify the promoter sequences of induced genes.

The promoter sequences were used as input to identify the TFs binding sites enriched in the promoters with the Plant Transcriptional Regulatory Map- Regulation Prediction (http://plantregmap.gao-lab.org/) using the default threshold p-value for binding site prediction (Jin et al., 2017). The regulation prediction identified the regulatory interactions and also the TFs binding sites which are enriched in the input promoter sequences. The enriched TFs were then cross referenced with the induced TFs in resistant genotype using the RNAseq data to find the TFs which are upregulated in resistant genotype post pathogen inoculation.

The enriched TFs which were upregulated in resistant genotype were further studied to see the downstream genes they are regulating using the regulatory map results and heat maps were generated based on the FPKM values of these genes in all four treatments: RM, RP, SM and SP. To further validate the binding site prediction, matrix of binding site in all the downstream induced genes was generated with the kp:Logo (http://kplogo.wi.mit.edu/submit.html) and they were compared with the database matrix for each TFs.

Further the cross-referenced TFs identified were grouped into primary and secondary TFs based on the TFs which were identified as downstream upregulated genes and are found co-expressed in the RNAseq data.

3.4 Results

3.4.1 Disease severity

The late blight severity in AG704 and Libertas cultivars was assessed by measuring the lesion diameter overtime. The initial symptoms were observed as brown coloured lesions on the inoculated leaves and stems at 3 dpi. The average lesion dimeter on leaflets was 69.25 mm in AG704 and in 35.1 in Libertas (Fig. 3.1A). The area under the disease progress curve (AUDPC), calculated using the lesion diameter, was 177.97 in AG704, which was significantly ($P < 0.01$) higher than in Libertas, 96.69 (Fig. 3.1C). The lesion diameter on stem was 20.1 and 7.5 mm for AG704 and Libertas, respectively (Fig. 3.1B). The AUDPC was significantly higher ($P < 0.01$) in AG704 (AUDPC = 57.18) compared to Libertas (AUDPC = 19.5) (Fig. 1D).

3.4.2 Resistance related induced (RRI) metabolites responsible for pathogen defense

Metabolites were extracted from inoculated foliage samples collected at 3 dpi and were analyzed based on LC-HRMS. A total of 3876 consistent monoisotopic masses were detected. The monoisotopic masses were further filtered to derive RR metabolites, which had higher fold change in resistant genotype compared to the susceptible. These were grouped into two categories, resistance related constitutive (RRC) and resistance related induced (RRI). The further studies were mainly focused on the RRI metabolites, as they are accumulated following the pathogen inoculation, with significantly higher abundance in the resistant genotype than in the susceptible.

A total of 281 RRC and 160 RRI metabolites were detected. In this study we have focused on induced metabolites and hence the RRC were ignored. The identities of these metabolites were further confirmed based on accurate mass error (AME <5ppm) and mass fragmentation pattern, which were matched with several databases, out of which 29 RRIs were accurately identified (Appendix I) and these belonged to different chemical groups. These metabolites were further classified into different metabolic pathways. The key phenylpropanoids, flavonoids and terpenoids accumulated in Libertas were: 10-Hydroxyloganin (FC= 9.43), secologanin (FC= 3.21), karanjin (FC= 3.46), solenolide (FC= 12.9), dihydroconiferyl alcohol (FC= 3.39), sinapoyl malate (FC= 3.01), steganacin (FC= 3.79), 7beta,12alpha-Dihydroxykaurenolide (FC= 3.83) and others (Table 3.1). Some conjugates were also identified, like oleanoic acid 3-O-glucuronide (FC= 9.18), vitexin

2"-O-beta-D-glucoside (FC= 4.89), quercetin 3-sophoroside (FC= 4.85), anhydroicaritin 3-rhamnosyl-(1->2)-rhamnoside (FC= 4.44), kaempferol 3-O-glucoside (FC= 3.33) and feruloylagmatine (FC= 12.53). Some of these metabolites are directly related to cell wall reinforcement and others are involved in pathogen suppression though antimicrobial properties. Other important RRI metabolites were related to lipid and fatty acid metabolism. A few metabolites were alkaloids and other toxins which can mainly function as a toxin for pathogen containment and others are further deposited in the cell wall.

3.4.3 Induced genes following pathogen invasion

Gene expression in leaves of resistant and susceptible genotypes, at 2 dpi with *P. infestans,* were obtained based on RNAseq. The *de novo* assembly of the reads was used to analyse the RNAseq outputs so that most transcripts can be detected. The analysis revealed 80,000 transcripts which were further filtered based on redundancy, and threshold length. The output was further filtered to obtain the differentially expressed genes (DEGs) and induced genes in resistant genotype were identified which were not induced in susceptible genotype (henceforth referred to as induced genes), for which the $\log_2$ fold change was calculated. A total of 611 induced genes (Appendix II) were obtained which were further categorized based on their biological functions. The Potato homologs of these *de novo* assembled transcripts were found using the command line BLAST and the NCBI transcript IDs were obtained from the NCBI GenBank.

A total of 52 receptor genes were found to be upregulated in the resistant genotype, of which lectins were predominant (Appendix II). The other induced genes were mostly related to the biosynthesis of phytoalexins or antimicrobial compounds such as lipids, fatty acids, phenylpropanoids, alkaloids, flavonoids and terpenoids, including hormone signalling (Fig. 3.2). Some of these genes were also involved in the biosynthesis of conjugated metabolites that reinforced the cell walls. Among the 611 induced genes, 74 TFs of different families were found, which generally regulate the pathogen defense response by regulating several metabolic pathways. 11 ethylene response factor (*ERF*), 8 *NAC*, 7 basic Leucine Rich Repeats (*bHLH*), 7 LOB Domain (*LBD*) Proteins, 5 *MYB* or *MYB* related genes were among the important TF family genes (Fig. 3.3B). To further map these genes to their biosynthetic genes we conducted a TF enrichment analysis (Fig. 3.3A).

3.4.4 Confirmation of RNAseq DEGs by qPCR

A total of 8 genes, out of 611 induced genes, were selected after filtering the RNAseq data to confirm the output. The primers were designed using the NCBI primer BLAST (Appendix III). The relative gene expression based on RNA seq was further confirmed based on qRT-PCR (Fig. 3.4).

3.4.5 Overview of transcription factor enrichment analysis

To better understand the regulatory framework of the induced genes, the cis elements in the induced genes were explored. To start with, 1000 base pair upstream promoter sequences of 611 induced genes were identified using an in-house script (Fig. 3.3A). These promoter sequences were then used as input for regulation prediction which also gives us TF enrichment. The TF enrichment analysis revealed 6465 regulations, between 284 TFs (from database) and 552 downstream genes (611 induced gene promoters) (Fig. 3.3A). Out of 284 distinct TFs, 134 were highly enriched and had binding sites in several induced gene promoters. These TFs were further categorized based on their family and it was evident that certain families of TFs like *MYB, bHLH, bZIP, Dof, TCP* and *C2H2* were predominant than others (Fig. 3.3B). Further, the binding site and the interaction scores for each TF with its downstream target genes were identified.

3.4.6 TFs regulating downstream metabolite biosynthetic genes

To explore the TFs regulating the defense response it was important to identify co-expressed TFs which show a better correlation of regulation between target genes and TFs. Accordingly, the 6465 regulations from the TF regulation prediction and enrichment study were cross referenced with the 74 induced TFs using the PGSC IDs and 14 TFs were identified. Among these *bHLH66, MYB61, NAC56, WRKY16*, *MYB like, ERF RAP2-3, MADS-box AGL15, NAC 2, HSF B-3, LOB 15, NAC 29, WRKY 23, E2F like* and *bZIP TGA10* TFs were found to regulate downstream induced genes. A total of 14 induced TFs, along with the binding motif numbers, PGSC protein IDs and the RRI Log2 FC are shown in Table 3.2.

The induced TFs had motif binding sites in several downstream induced genes. *MYB61* had binding site in 42 downstream genes, including 7 TFs, whereas *WRKY16* had binding sites in 26 downstream induced genes, including 2 TFs. The *ERF* and *MADS-box AGL15* had binding sites in 21 downstream genes each, of which 4 were TFs. The *NAC56, bHLH66* and *MYB like* TFs had binding sites in 20, 19 and 16 induced downstream genes, respectively, and they regulated 7 other TFs. Other TFs *viz, NAC2, HSF B-3, LOB15, NAC29, WRKY23, E2F like, bZIP TGA10* had less than 10 target genes. The TFs identified were further classified into primary and secondary TFs, which directly or indirectly regulate the downstream biosynthetic genes (Fig. 3.5A).

The transcriptional reprogramming due to these primary TFs upregulate the defense related genes only after pathogen attack, as revealed in a heat map of all the downstream genes regulated with respect to each treatment in RNAseq (Fig. 3.6, 3.7 & 3.8). To explore further, the binding motif matrix for each of these TFs (Table 3.2), and the binding site of each gene promoter was explored. It was evident that the binding site matrices match each other to a great degree (Fig. 3.9).

3.4.7 Downstream TFs and their relationship with biosynthetic genes

The membrane located receptor genes perceive the invading pathogen signals and trigger the TFs, which mount the transcriptional reprogramming by regulating other TFs, and these can have major or minor impacts based on the number or genes they regulate. Hence, a two-tier transcriptional network was mapped (Fig. 3.5 A). The key TFs found in the enrichment study had regulatory binding motifs in other TF promoter regions. The 14 TFs cross referenced were grouped into primary TFs whereas the target TFs identified as downstream induced genes were grouped into secondary TFs (Fig 3.5 A). These downstream TFs were mainly from *NAC, LBD, MYB, bZIP* and *WRKY* family. These TFs, based on their induced gene Log2 FC, were grouped according to their master switching of TFs (Fig 3.5 B &C)

3.4.8 Association of TFs with resistance genes and metabolites

It is also important to explore the relationship between the disease resistance response triggered by the TFs via induced genes. The TFs not only regulate secondary TFs but also other downstream genes which not only encode resistance proteins but also biosynthesize resistance metabolites

(Table 3.3). After correlating the TFs with downstream genes and metabolites it appeared that *MYB61* and *LBD* affect the phytoalexin production and cell wall reinforcement, whereas *WRKY16* and *ERF* were mostly seen affecting the pathogenesis related protein production. *MYB61* regulates the phenylpropanoid, loganin, flavonolignans and lignan biosynthesis. The genes that *MYB61* regulate are scopoletin glucosyltransferase (FC= 4.25), dirigent protein (FC= 3.37), 7-deoxyloganetic acid glucosyltransferase (FC= 1.85), 7-deoxyloganetin glucosyltransferase (FC= 2.25) and trans-cinnamate 4-monooxygenase (FC= 2.02) (Table 3.3). Hence, *MYB61* seems to have a larger impact on disease resistance. *MYB like* and *LBD* TFs seem to affect the mevalonate pathway via hydroxy methyl glutaryl-CoA synthase (FC= 3.57) and 3-hydroxy-3-methylglutaryl-coenzyme A reductase (FC= 4.00), which is directly linked to terpenoid biosynthesis. *LBD* is also linked to lipid biosynthesis by regulating the beta-amyrin 28-monooxygenase (FC= 3.13) and 12-oxophytodienoate reductase 1 (FC= 5.18) genes (Table 3.3). *WRKY16* and *ERF* on the other hand mostly affect the hypersensitive response and pathogenesis related proteins.

3.5 Discussion

Plant genotypes varying in disease resistance may have several common mechanisms of resistance though the distinct ones make them more resistant. The resistance is mainly due to antimicrobial metabolites and proteins, or their deposition to reinforce the cell walls around infected area to contain the pathogen to initial infection area. The resistance response to a disease may be due to several metabolites that are controlled by a network of metabolic pathways regulated by TFs (Meraj et al., 2020).

In this study, several resistance induced metabolites have been implicated to be responsible in suppressing the pathogen, and these also have been previously reported from other plant-pathogen interaction studies, including potato-*Phytophthora* interaction (Pushpa et al., 2013; Yogendra et al., 2014; Yogendra et al., 2015b). These metabolites have been proved to be responsible for cell wall fortification and as well as antifungal metabolites (Yogendra et al., 2015a; Kage et al., 2017). The induced metabolite biosynthetic genes on the other hand are triggered by the TFs which play a vital role in transcriptional reprogramming during pathogen attack in plants. In this study, the overall picture of transcriptional reprogramming, which happens downstream to the pathogen signal perception and transduced through MAPK or CDPK was explored. The aim of this study

was to narrow down to the master TFs which play a major role in reprogramming the defense response (Fig. 3.10).

3.5.1 Metabolo-transcriptomic study articulate the disease response

The semi comprehensive metabolomics revealed several metabolites that are more accumulated in the resistant genotype, Libertas, than in the susceptible genotype, AG704. These metabolites mainly belong to the flavonoid, fatty acid and lipid, terpenoid, alkaloid and phenylpropanoid pathways. The transcriptomic studies also reiterate the metabolomic narrative showing the genes which are mainly responsible for the biosynthesis of the metabolites which seem to be accumulated post pathogen inoculation. In some of the previous studies in potato and other plant species a lot of these genes and metabolites have been reported as either constitutively synthesized or induced following pathogen invasion (Pushpa et al., 2013; Yogendra et al., 2014; Yogendra et al., 2015b; Dhokane et al., 2016; Yogendra and Kushalappa, 2016; Karre et al., 2017). However, the TFs which directly induce these genes and indirectly the metabolites have never been studied by analyzing the promoter regions of induced genes. Plant has to switch from regular mode to plant defense mode, through transcriptional reprogramming to mount defense response, as these processes are energy dependent, and hence it is very important to study the inducing factors like TFs (Macho and Zipfel, 2014) (Fig. 3.10). Based on the present study we present an overview of the defense response which involves the transcriptional reprogramming and starts with the pathogen perception which acts a trigger. (Fig. 3.10). This trigger is further trickled down to the nucleus through the MAPK or CDPK and the transcriptional reprogramming happens through the primary and secondary TFs which either activate or repress downstream pathway genes and leads to mounting the defense response (Fig. 3.10)

3.5.2 Transcriptional reprogramming post infection

Arabidopsis has 1600 different TFs, whereas potato has ~2400 TFs, grouped into different families. Most of these TFs have a specific role and regulate a specific signal. They are useful for fine-tuning the response to a particular stimulus along with other regulatory mechanisms like small RNA, histone modifications and post translational modifications (Garner et al., 2016). In this

study, 74 TFs were found that were differentially upregulated in resistant genotype post pathogen inoculation. Out of these 30 TFs were further mapped and 14 TFs were grouped into primary upstream TFs and 16 downstream secondary TFs (Fig 3.5A). It may be possible that the other TFs are involved in downregulation of genes. A similar two tired transcriptional programming of *NAC* and *MYB* TFs is also found in secondary cell wall biogenesis (Nakano et al., 2015). It is also important to note that the primary TFs identified in this study are from different families, which entails cumulative effects of these primary regulators that may play a synergistic role in defense response.

3.5.3 Synergistic effects of primary and secondary transcription factors

In a recent study, several *MYB* and *bHLH* TFs were identified as primary TFs which regulate the anthocyanin biosynthesis through co-expression studies (Zhang et al., 2017). Also, *NAC* and *MYB* are involved in secondary cell wall biosynthesis in specialized cells like xylem vessels (Nakano et al., 2015). Hence, disease resistance which requires several pathways to be activated and several others to be suppressed should involve regulation through multiple factors.

The primary TFs like *MYB61* in *Arabidopsis* have shown to be involved in secondary cell wall biogenesis and terpene metabolism (Voiniciuc et al., 2015; Matías-Hernández et al., 2017). Also, orthologues of *WRKY16* are shown to be involved in dual function of regulating pathogenesis related proteins and sinapoyl metabolites (Hussain et al., 2018b; Hussain et al., 2018a). In tomato *ERF RAP2-3* closest relative is shown to be involved in root nematode defense and crosstalk between defense hormones (Zhao et al., 2018). A similar synergistic approach is shown in the network analysis between several TFs and metabolic genes for secondary cell wall biosynthesis which mainly focuses on abiotic stress (Taylor-Teeples et al., 2015). However, similar study is still lacking which focuses on biotic stresses.

3.5.4 Network studies in potato disease resistance

Transcriptional reprogramming is a dynamic event and has a complex network involving several signaling mechanisms. A similar network map was developed for tomato against *Pseudomonas syringae* which mainly focused on the effector triggered immunity, encompassing the early

response to pathogen signal (Lewis et al., 2015). However, no other study mentions a similar network analysis focusing on the quantitative disease response in plants. A network model was developed based on *Arabidopsis* network model, which shows a crosstalk between salicylic acid and ethylene signaling but it was with Potato Virus Y (PVY) transcriptomics data as reference (Ramšak et al., 2018). It can be articulated that several differences lie between different pathogen signaling pathways and it is important to identify them for better understanding of the resistance mechanisms in plants. In previous studies on potato-*Phytophthora* interaction several TFs were identified which regulate a few genes or individual pathways (Yogendra et al., 2015a; Yogendra et al., 2017). However, the present study focusses on the TFs not related to a single pathway, but on co-expressed genes and respective TFs through the cis-regulatory elements (Table 3.2).

Specific studies relating to the jasmonate responsive TFs have shown how different *WRKY, ERF, bHLH* and *MYB* TFs regulate the JA signaling and glucosinolate metabolism (Zhou and Memelink, 2016). Some other studies show the plant hormone homeostasis during the defense response, which states that TFs play a crucial role in regulating the hormone signaling which makes hormones as secondary regulators (Zhou and Memelink, 2016). Specific analysis to find large scale genome binding targets of TFs are limited when it comes to plant defense and hence more evidences are necessary to find the key transcriptional regulators (Sun et al., 2015; Birkenbihl et al., 2017a). It is also important to note that this study is at a fixed time point and there may be other targets of these TFs which may be activated before or after 48 hpi.

Several other co-expression studies have shown a global overview of the relationship between the transcriptional regulators and cis-elements (Wong et al., 2018; Gómez-Cano et al., 2019). It is also seen that some hormones activate specific TFs, but exact mapping of factors which regulate *in vivo* hormone signaling is still elusive (Wiesel et al., 2015). On the other hand, in this study an attempt was made to find the large impact TFs which could be further useful for the identification of *in vivo* targets by chromatin immunoprecipitation (ChIP) analysis. It also takes into account the metabolic difference which is the phenotypic effect of the transcriptional reprogramming, and metabolites are considered closest to the actual phenotype (Kushalappa and Gunnaiah, 2013; Kushalappa et al., 2016). No reports of *in vivo* ChIP analysis have been reported in potato and it is also missing for several other plant species except for the model plants like *Arabidopsis* (Liu et al., 2015). It is necessary to validate these TFs *in vivo* to further understand their role in resistance,

which can be useful to formulate a bigger project like ENCODE (Encyclopedia of DNA Elements) in plants and study like the present one will be a steppingstone for the same (Birkenbihl et al., 2017b). Though there is an enormous data available on the disease resistance in terms of genes and metabolites, still it is not possible to understand the master switches of these components (Zhang et al., 2019).

3.6 Conclusion

Plant diseases play a crucial role in world food production and economic stability in general. Hence, managing and containing the deadly plant diseases like late blight is of paramount importance. Though a lot of research has been done in this area, it is still unclear why certain genotypes are resistant, while others are susceptible to a particular disease. Here an attempt is made to understand the transcriptional differences in resistant and susceptible genotypes inoculated with pathogens and characterized the primary and secondary TFs based on their expression and transcription binding sites in the promoters, eventually to mount an effective defense based on its repertoire of induced genes.

Table 3.1 Resistance related induced metabolites with fold change in abundance identified in potato following inoculation with *P. infestans*

	Metabolites	Fold Change	Observed Mass (Da)	Actual Mass (Da)	AME
Terpenoids					
	Erythroxanthin sulfate	26.8*	678.3613	678.359	3.4
	(-)-Solenolide A	12.9*	554.2279	554.2283	-0.72
	Cymarin	12.08**	548.2981	548.2985	-0.73
	Cerberin	11.64**	576.3309	576.3298	1.91
	10-Hydroxyloganin	9.43***	406.1483	406.1475	1.88
	Oleanoic acid 3-O-glucuronide	9.18*	632.3928	632.3924	0.63
	Kanokoside D	5.51***	624.2629	624.2629	-0.02
	Vernoflexuoside	4.55*	408.1795	408.1784	2.66
	Officinalisnin	4.18*	920.4962	920.4981	-2.07
	Dehydrovomifoliol	3.89*	222.1257	222.1256	0.63
	7beta,12alpha-Dihydroxykaurenolide	3.83*	332.1982	332.1988	-1.91
	Patrinoside	3.7**	462.2119	462.2101	3.83
	Secologanin	3.21*	388.1383	388.1369	3.73
Flavonoids					
	8-Prenylnaringenin	9.01**	340.1317	340.1311	1.67
	Vitexin 2"-O-beta-D-glucoside	4.89*	594.159	594.1585	0.84
	Quercetin 3-sophoroside	4.85**	626.1493	626.1483	1.64
	Anhydroicaritin 3-rhamnosyl-(1->2)-rhamnoside	4.44**	660.245	660.2418	4.85
	Karanjin	3.46***	292.0748	292.0736	4.09
	Kaempferol 3-O-glucoside	3.33*	448.1012	448.1006	1.37
	Lupinisoflavone J	3.04***	438.167	438.1679	-2.08
Alkaloids					

	Alangicine	7.34*	480.2635	480.2624	2.3
	Serratine	4.64**	279.1845	279.1834	3.8
	Solanine	2.66*	867.4995	867.498	1.78
Phenylpropanoid					
	Feruloylagmatine	12.53***	306.1702	306.1692	3.43
	Robustaol A	12.18***	474.1879	474.189	-2.37
	Dihydroconiferyl alcohol	3.39*	182.0941	182.0943	-1.02
	Sinapoyl malate	3.01*	340.0805	340.0794	3.3
	Steganacin	3.79*	456.141	456.142	-2.09
Fatty acids and lipids					
	N-heptanoyl-homoserine lactone	7.34*	213.1367	213.1365	1
	PI(17:2(9Z,12Z)/0:0)	4.33*	582.2804	582.2805	-0.2
	(9Z)-(7S,8S)-Dihydroxyoctadecenoic acid	3.84*	314.2463	314.2457	2.05
	6,9,12,15-octadecatetraenoic acid	3.79**	276.21	276.2089	4.17
	2-pentadecenoic acid	3.2*	240.2097	240.2089	3.23

- Observed Mass: To the observed mass one H mass was added because the LC/MS analysis was done in negative ionization mode.
- Fold change was calculated based on relative intensity of metabolites: FC= ((RP/RM)/(SP/SM)), FC=fold change, RP= resistant pathogen inoculated, RM- resistant mock inoculated, SP= susceptible pathogen inoculated and SM- susceptible mock inoculated.
- *t test significance at $P< 0.05$, ** t test significance at $P< 0.01$ and *** t test significance at $P <0.0001$.
- Da=Daltons
- AME = accurate mass error (ppm); AME is calculated by formula ((Observed m/z - actual m/z) / observed m/z)) $*10^6$

Table 3.2 Transcription Factor Enrichment analysis.

Transcription Factor	PGSC Protein ID	Log2 FC (RRI)	Binding Motif ID	No. of target genes
MYB 61	PGSC0003DMP400027651	5.43	MP00134	42
WRKY 16	PGSC0003DMP400054355	2.68	MP00531	26
ERF RAP2-3	PGSC0003DMP400029373	3.74	MP00109	21
MADS-box AGL15	PGSC0003DMP400011976	5.67	MP00508	21
NAC 52	PGSC0003DMP400033522	5.88	MP00361	20
bHLH66	PGSC0003DMP400019655	1.89	MP00659	19
MYB like	PGSC0003DMP400001575	2.09	MP00160	16
NAC 2	PGSC0003DMP400016317	1.65	MP00121	9
HSF B-3	PGSC0003DMP400047457	3.83	MP00305	8
LOB 15	PGSC0003DMP400053002	12.16	MP00283	8
NAC 29	PGSC0003DMP400005111	4.93	MP00221	6
WRKY 23	PGSC0003DMP400015928	1.65	MP00069	5
E2F like	PGSC0003DMP400042307	2.16	MP00397	5
bZIP TGA10	PGSC0003DMP400040959	3.92	MP00491	2

A list of the primary TFs found after correlating the induced R genes and the promoter elements of the induced genes. These form the first tier in the two-tier transcriptional reprogramming map. The PGSC protein IDs and the log2 fold change along the promoter binding motif IDs are given in the table for each TF. The table also mentions the number of downstream genes each transcription factor regulates.

Table 3.3 Transcription factors identified in potato cultivars following inoculation with mock and *P. infestans*, and the metabolites and metabolic pathways they regulate

	Gene	FC	PGSC Transcript ID	Uniprot ID	Metabolite	Pathway/ Function
MYB like						
	Hydroxymethylglutaryl-CoA synthase	3.57	PGSC0003DMT400032619	P54873	7beta,12alpha-Dihydroxykaurenolide	Mevalonate pathway
	BAHD acyltransferase DCR	6.93	PGSC0003DMT400069898	Q9FF86	6,9,12,15-octadecatetraenoic acid	Cutin Biosynthetic pathway
	CDP-diacylglycerol--serine O-phosphatidyltransferase 1	2.23	PGSC0003DMT400078518	F4HXY7	PI(17:2(9Z,12Z)/0:0)	Phospholipid biosynthesis
	Thaumatin-like protein 1	3.72	PGSC0003DMT400058117	A0A1P8B554	-	PR Protein
	Anthranilate N-benzoyltransferase protein 1	5.66	PGSC0003DMT400080508	O24645	ND	Benzoylanthranilate biosynthesis
	Senescence-specific cysteine protease SAG39	11.69	PGSC0003DMT400026448	Q7XWK5	-	Hypersensitive Response
	Disease resistance protein RPS5	2.05	PGSC0003DMT400055296	O64973	-	Pathogenesis Related Protein
	Snakin-2	1.90	PGSC0003DMT400040298	Q93X17	-	Pathogenesis Related Protein
bHLH66						
	Thaumatin-like protein	3.15	PGSC0003DMT400061140	P50699	-	Pathogenesis Related Protein

MYB 61						
	Scopoletin glucosyltransferase	4.25	PGSC0003DMT400066570	Q9AT54	Dihydroconiferyl alcohol	Phenylpropanoid
	L-ascorbate peroxidase 1 cytosolic	1.68	PGSC0003DMT400077286	Q05431	-	Detoxification
	Dirigent protein 23	3.37	PGSC0003DMT400004111	Q84TH6	ND	Flavonolignans
	7-deoxyloganetic acid glucosyltransferase	1.85	PGSC0003DMT400059654	U3U992	10-Hydroxyloganin	Loganin Biosynthesis
	Calmodulin-like protein 3	2.64	PGSC0003DMT400018016	Q9SRR7	-	Hypersensitive Response
	IQ domain-containing protein IQM2	3.35	PGSC0003DMT400080788	Q9LHN9	-	Hypersensitive Response
	Premnaspirodiene oxygenase	5.06	PGSC0003DMT400028546	A6YIH8	ND	Antifungal Compound
	7-deoxyloganetin glucosyltransferase	2.25	PGSC0003DMT400032758	F8WKW1	Secologanin	Loganin Biosynthesis
	Trans-cinnamate 4-monooxygenase	2.02	PGSC0003DMT400078290	P48522	ND	Lignan Biosynthesis
NAC 56						
	Cytochrome P450 CYP72A219	2.35	PGSC0003DMT400027316	H2DH21	ND	Ginsenoside biosynthesis
	Phenylpropanoid glucosyltransferase 1	3.65	PGSC0003DMT400066574	Q9AT54	Sinapoyl malate	Phenylpropanoid
WRKY 16						

	Disease resistance protein RPM1	1.97	PGSC0003DMT400047579	Q39214	-	Pathogenesis Related Protein
	Monothiol glutaredoxin-S3	2.57	PGSC0003DMT400079099	O23421	-	Hypersensitive Response
	Osmotin-like protein OSML13	1.74	PGSC0003DMT400007870	P50701	-	Pathogenesis Related Protein
	Peroxidase 4	11.20	PGSC0003DMT400031195	A7NY33	-	Hypersensitive Response
ERF						
	Disease resistance protein RPS5	2.05	PGSC0003DMT400055296	O64973	-	Pathogenesis Related Protein
	Pathogenesis-related leaf protein 4	2.10	PGSC0003DMT400005136	Q04108	-	Pathogenesis Related Protein
HSF						
	Pathogen-related protein	2.94	PGSC0003DMT400063921	P16273	-	Pathogenesis Related Protein
LBD						
	3-hydroxy-3-methylglutaryl-coenzyme A reductase 2	4.00	PGSC0003DMT400008902	P48022	7beta,12alpha-Dihydroxykaurenolide/Solanine	mevalonate pathway
	Inositol oxygenase 4	2.29	PGSC0003DMT400000780	Q8H1S0	ND	Cellulose Biosynthesis
	Beta-amyrin 28-monooxygenase	3.13	PGSC0003DMT400008432	Q2MJ20	Oleanoic acid 3-O-glucuronide	Oleanolate biosynthesis
	12-oxophytodienoate reductase 1	5.18	PGSC0003DMT400048344	Q9XG54	(9Z)-(7S,8S)-Dihydroxyoctadecenoic acid	Lipid biosynthesis

MIKC_MADS						
	Sphingosine kinase 1	2.89	PGSC0003DMT400012976	Q8L7L1	2-pentadecenoic acid	Lipid biosynthesis
	Protein STRICTOSIDINE SYNTHASE-LIKE 6	1.69	PGSC0003DMT400003995	Q9SD05	ND	Camptothecin biosynthesis
	Kirola	9.64	PGSC0003DMT400041449	P85524	-	Pathogenesis Related Protein

The table shows the relation between the transcription factor, genes regulated by these transcription factors and the disease related activity. The details of metabolites are given in Table 3.1.

ND- Not Detected

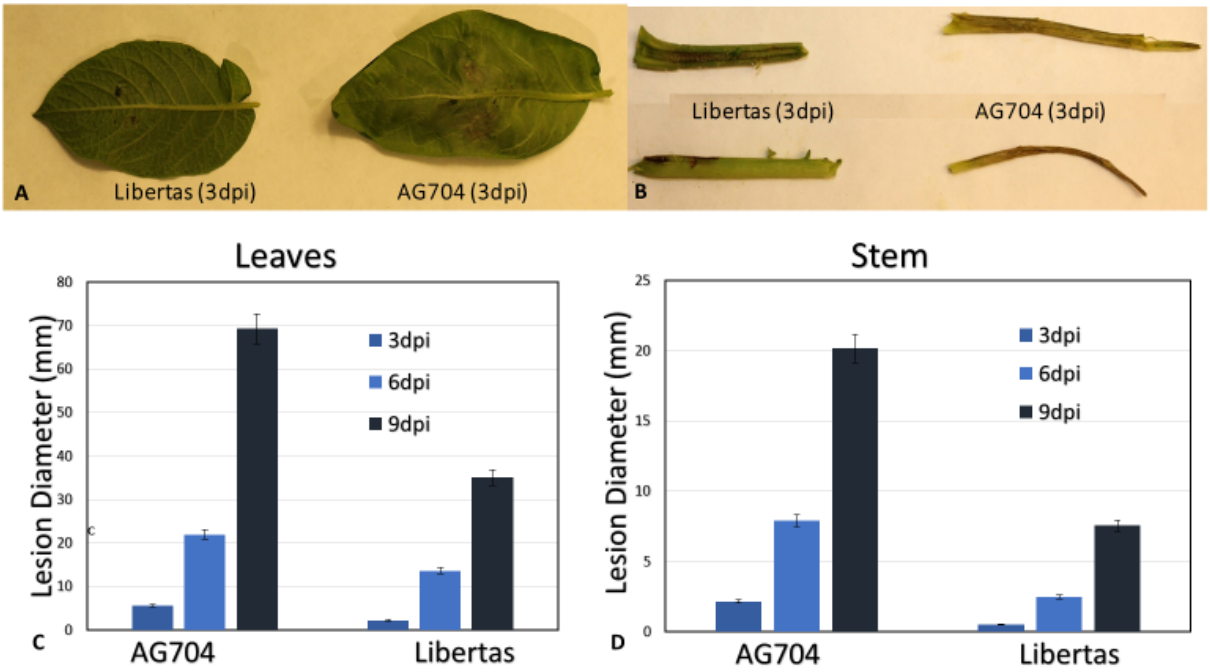


Fig 3.1 Disease symptoms and the lesion diameter diseased at different dpi.

A) Shows the 9 days post inoculation leaf of AG704 and Libertas where a distinct difference in disease severity increase over time. The pathogen is well contained in Libertas whereas it has drastically progressed in the AG704 leaves. B) Shows the 9 days post inoculation stem samples of AG704 and Libertas where similar to leaves the pathogen progression can be seen. C and D) The lesion diameter on stem was recorded and the diameter was 20.1 mm and 7.5 mm for AG704 and Libertas respectively. The AUDPC calculated in stem inoculations was significantly high ($P < 0.01$) in AG704 (AUDPC = 57.18) compared to Libertas (AUDPC = 19.5).

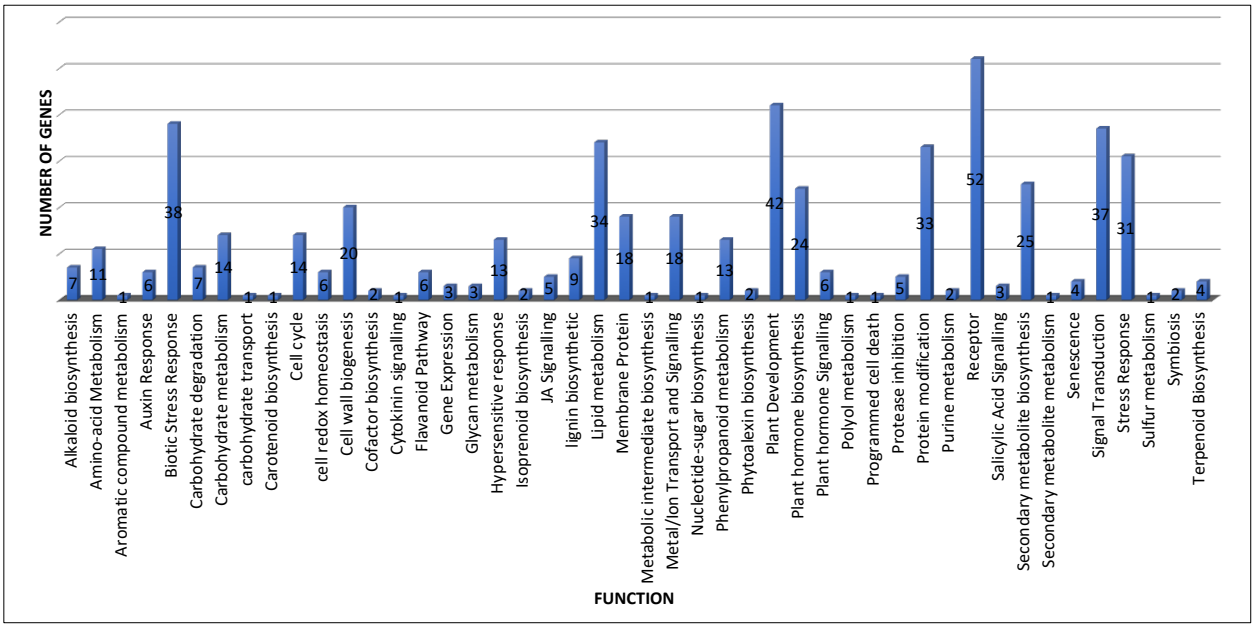


Fig. 3.2 Induced genes identified and their biological functions.

The bar graph shows induced genes grouped based on their biological functions. The distribution was based on the manual curation of each gene based on the data available on the Uniprot database as well as pathway analysis of the induced genes. It is seen that most of the genes were distributed in the categories of receptor, biotic stress response, plant development, signal transduction, lipid metabolism, protein modification and general stress response. Several other specific pathway genes were also found like cell wall biogenesis, lignin biosynthesis, hypersensitive response which are important in case of disease resistance.

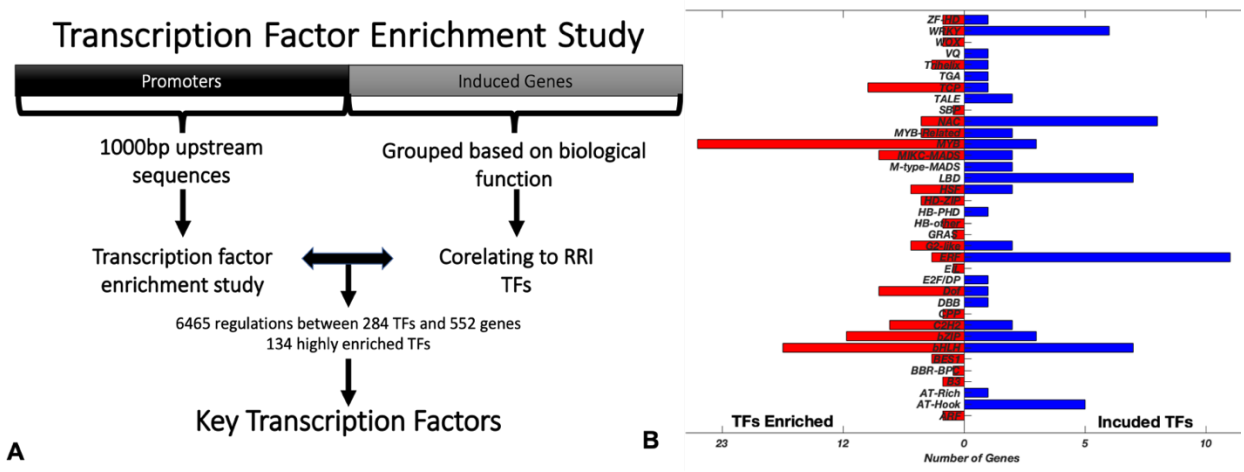


Fig 3.3 Distribution of transcription factors based on their families.

A) An overview of the transcription factor enrichment study. B) The transcription factors identified found among the resistance related induced R and highest number of transcription factors are *ERF, NAC, bHLH* and *WRKY*. shows the distribution of transcription factors found in the transcription factor enrichment study, which was based on the promoter regions (-1000bp) of the RRI genes. Here it can be seen that *MYB, bHLH* and *Dof* family transcription factors were highly enriched. (RM- resistance mock inoculations, RP- resistance pathogen inoculated, SM- susceptible mock inoculated, SP- susceptible pathogen inoculated).

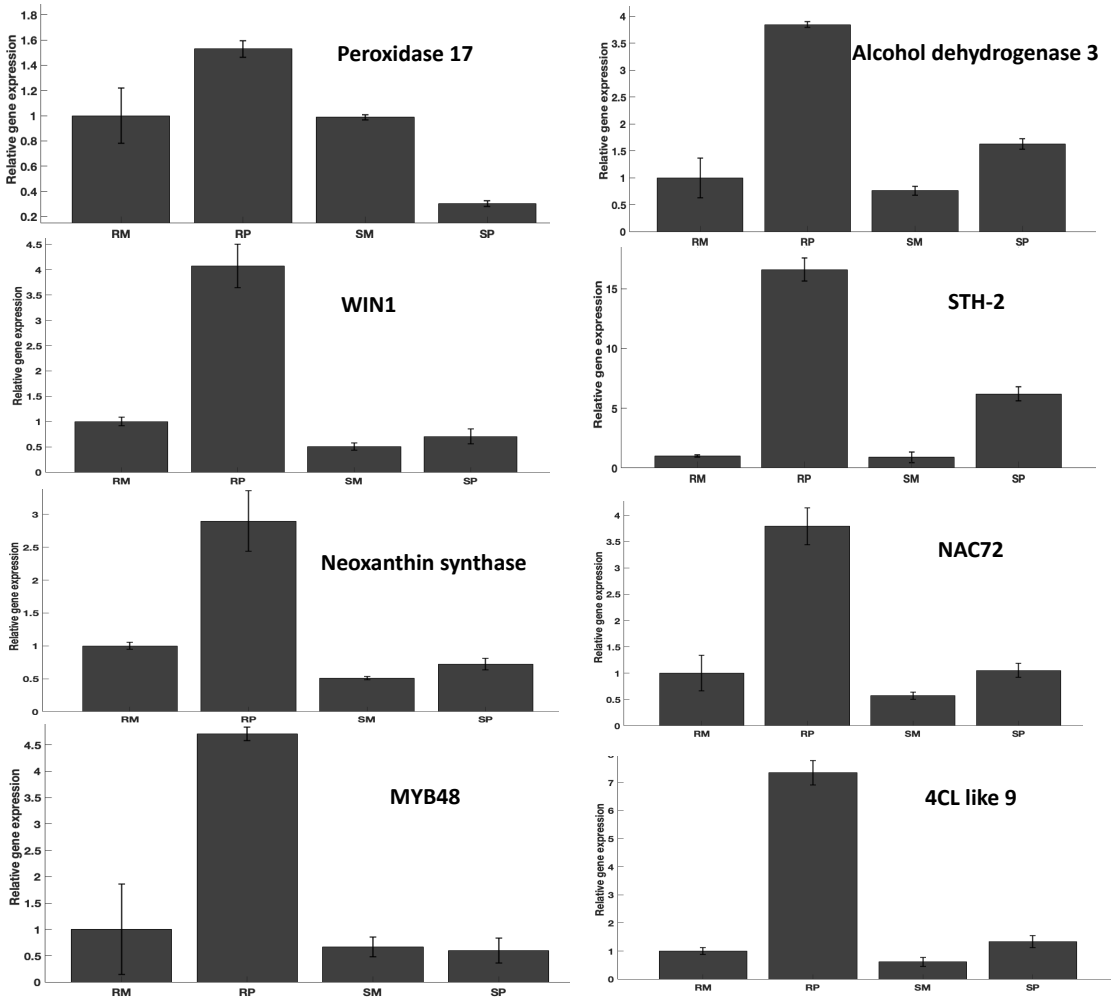


Fig. 3.4 qPCR analysis of selected genes from RNAseq data.

PCR of selected genes for the confirmation of gene expressions quantified based on RNA seq data, of potato genotypes inoculated with mock and *P. infestans*. .The figure shows the relative gene expression data obtained by qPCR for selected 8 genes from the induced R gene list. The primer sequences used for qPCR are given in Appendix III. The qPCR results correspond with the RNAseq FPKM values for the selected genes.

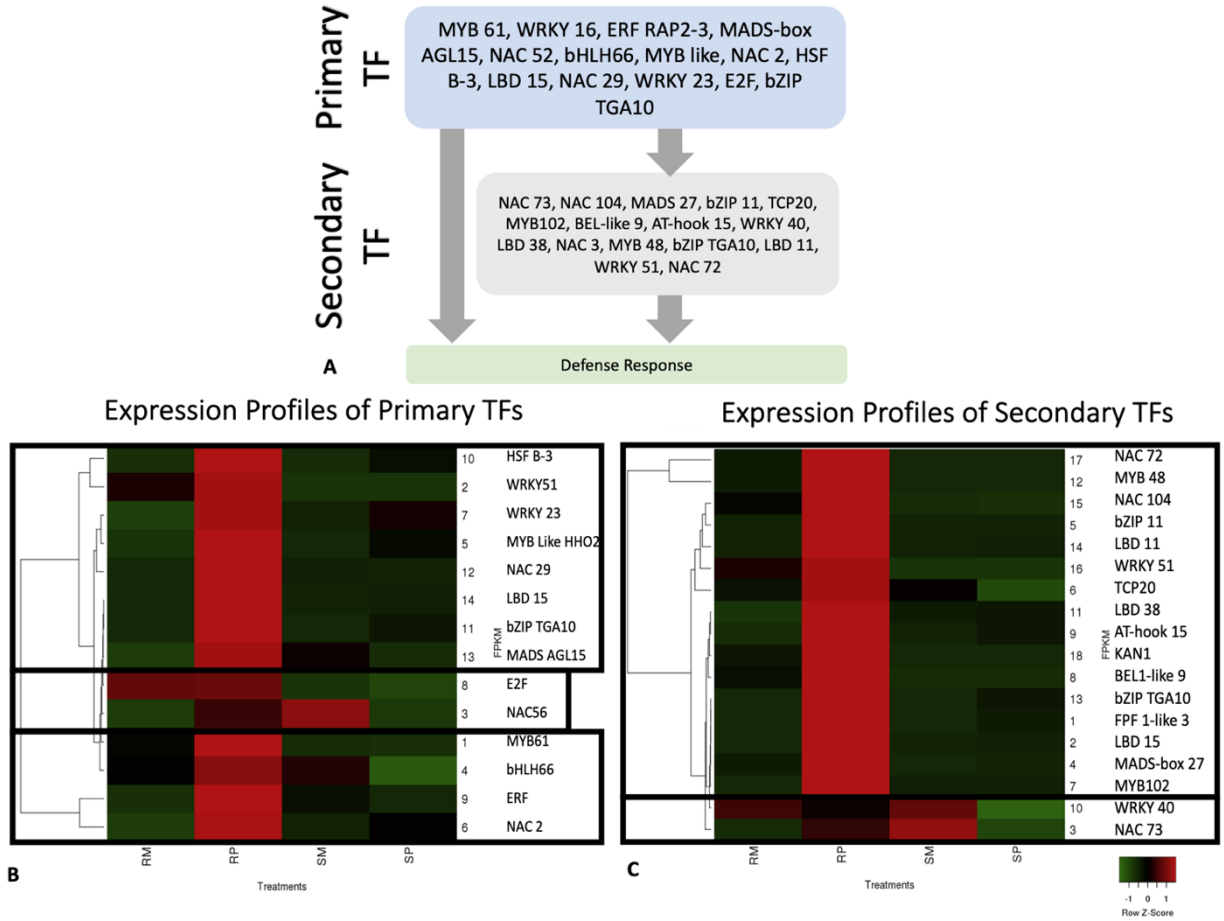


Fig. 3.5 Two-tier transcriptional reprogramming of genes identified by RNAseq and TF enrichment analysis.

A) Shows the characterization of induced TFs into primary and secondary. These TFs either directly or indirectly influence the disease response in a two-tier way. B) Shows the heatmap of primary transcription expression across different treatments (RM- resistant mock, RP- resistant pathogen inoculated, SM- susceptible mock, SP- susceptible pathogen inoculated). All the transcription factors show that they are upregulated in RP and have no change in susceptible across all treatments. However, the map shows that E2F has no drastic change in RP when compared to RM. Whereas, NAC56 shows it has no much change in RP when compared to RM, and it has been downregulated in SP when compared to SM. C) Shows the heatmap of secondary transcription expression across different treatments, that RM, RP, SM and SP. All the secondary TFs show higher expression in RP when compared to other treatments. However, WRKY 40 and NAC 73 have been downregulated in susceptible genotype upon pathogen inoculation.

Transcription Factor	PGSC Protein ID
Myb 61	PGSC0003DMP400027651

Transcription Factor	PGSC Protein ID
WRKY 16	PGSC0003DMP400054355

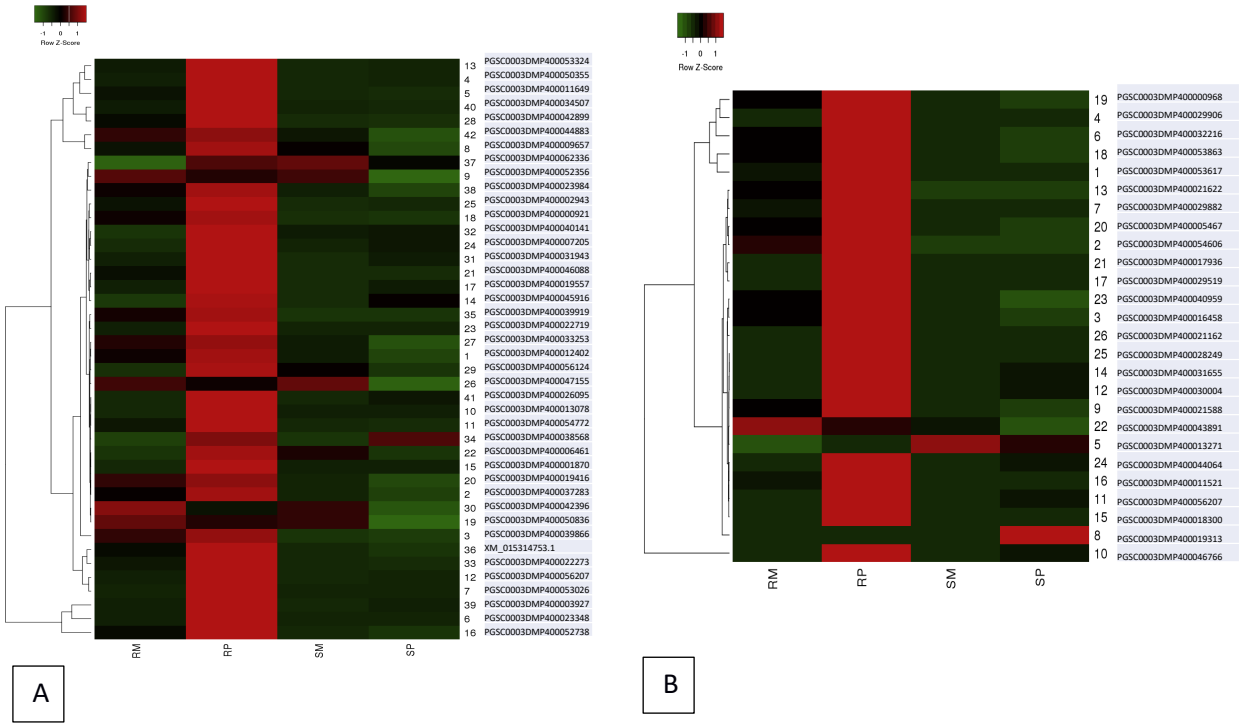


Fig. 3.6 Heat map of downstream genes regulated by *MYB 61* and *WRKY 16*.

Shows the expression levels in the form of heat maps of the genes regulated by *MYB61* (A) and *WRKY16* (B). We can see that some genes are upregulated in the susceptible mock inoculated samples which shows that these genes are downregulated upon pathogen attack however, they are either upregulated or their expression levels does not change post inoculation in resistant. (RM- resistance mock inoculations, RP- resistance pathogen inoculated, SM- susceptible mock inoculated, SP- susceptible pathogen inoculated).

Transcription Factor	PGSC Protein ID
NAC 52	PGSC0003DMP400033522

Transcription Factor	PGSC Protein ID
bHLH66	PGSC0003DMP400019655

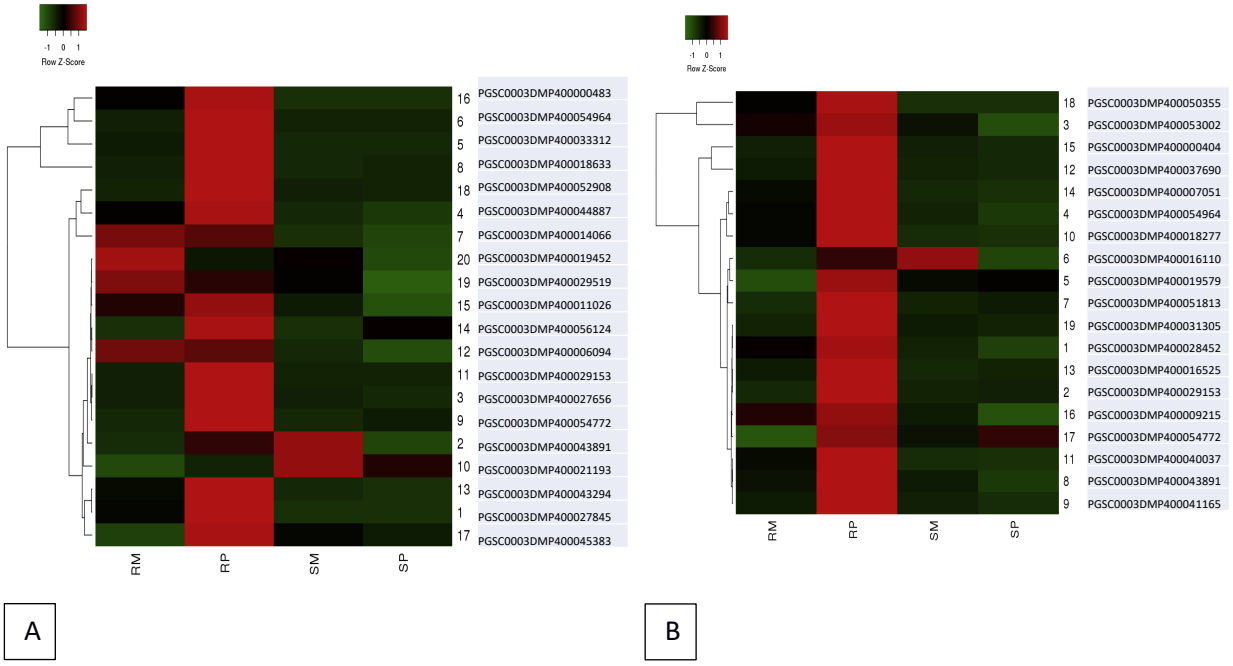


Fig. 3.7 Heat map of downstream genes regulated by *NAC 52* and *bHLH 66*.

Shows the expression levels in the form of heat maps of the genes regulated by *NAC52* (A) and *bHLH66* (B).

Transcription Factor	PGSC Protein ID
DNA binding protein HHO2	PGSC0003DMP400001575

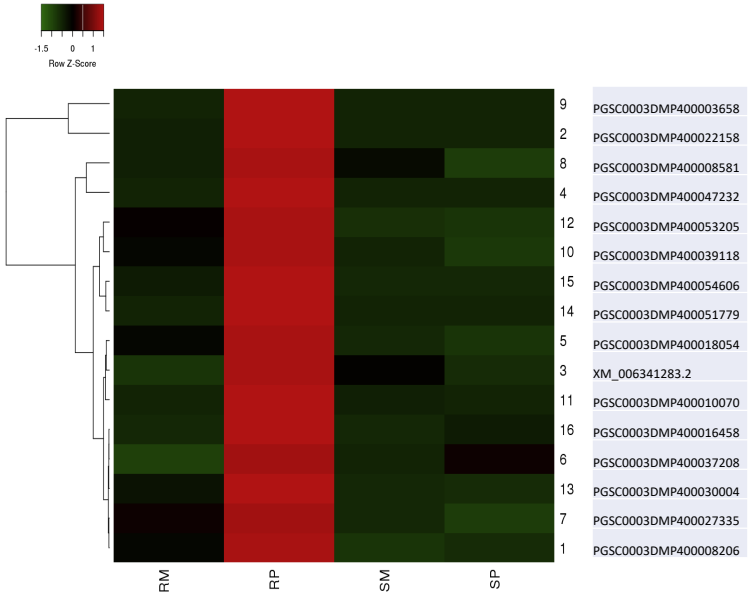


Fig. 3.8 Heat map of downstream genes regulated by *MYB* like transcription factor.

Shows the expression levels in the form of heat maps of the genes regulated by *MYB like* which is annotated as DNA binding HHO2 in the RNA-seq data.

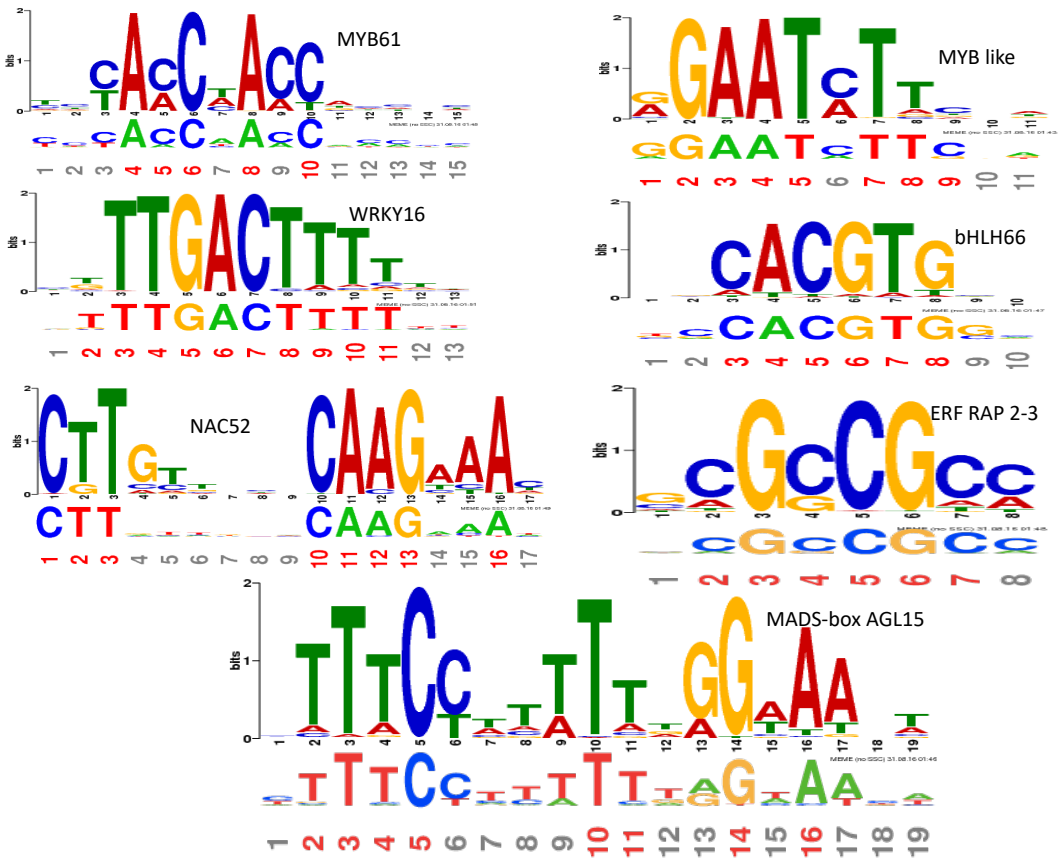


Fig. 3.9 Comparison of key transcription factor binding sites from database and the target gene promoter sequences.

This figure represents the TF binding motif for the key TFs (targets >15), compared with the binding motif extracted from the promoter regions of each of its induced R target genes. The top matrix for every TF is obtained according to the binding motif number for that particular TF and below it is the matrix representing the induced genes target binding sites.

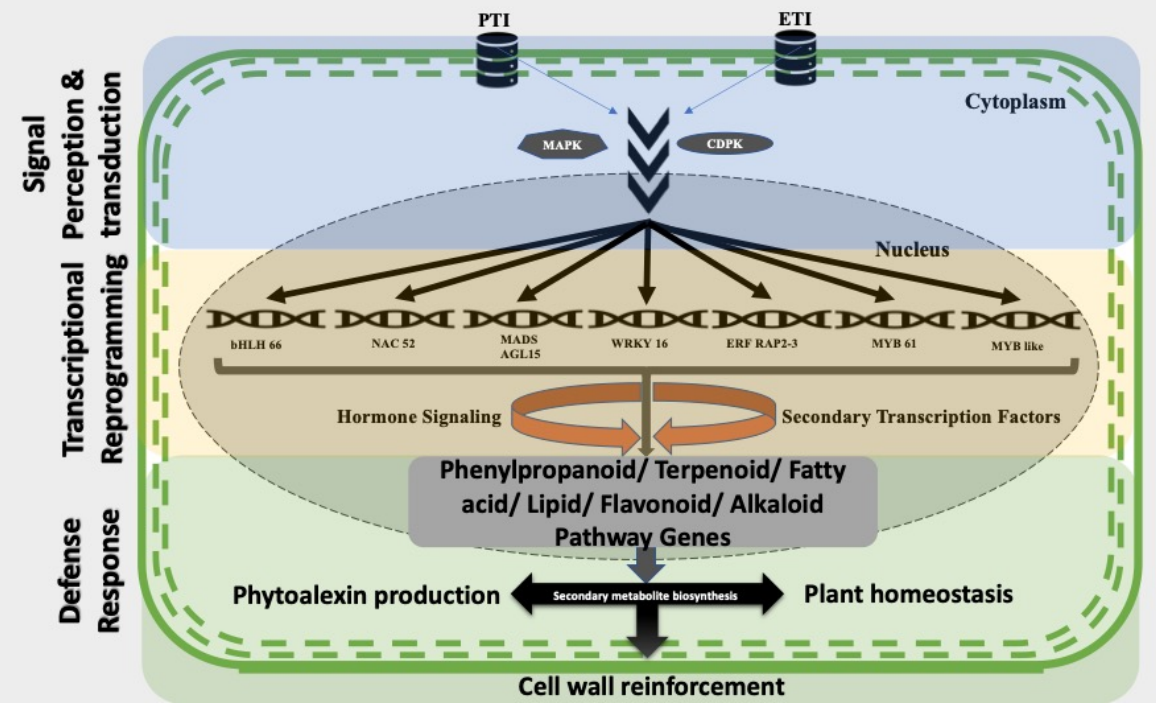


Fig. 3.10 Overview of transcriptional reprogramming to enhance disease resistance.

The figure shows the overview of the defense response orchestrated via the transcriptional reprogramming. The first division shows the pathogen perception through the PAMP triggered immunity (PTI) or effector triggered immunity (ETI) receptors, which is further trickled down to the nucleus through the MAPK or CDPK. The central layer is where the transcriptional reprogramming happens where the primary and secondary TFs activate or repress downstream pathway genes, which further lead to the defense response by the suitable metabolic changes. The primary and secondary TFs also regulate the hormones which in turn regulate certain metabolites.

Reference list

Birch PRJ, Bryan G, Fenton B, Gilroy EM, Hein I, Jones JT, Prashar A, Taylor MA, Torrance L, Toth IK (2012) Crops that feed the world 8: Potato: Are the trends of increased global production sustainable? Food Secur **4**: 477–508

Birkenbihl RP, Kracher B, Roccaro M, Somssich IE (2017a) Induced Genome-Wide Binding of Three Arabidopsis WRKY Transcription Factors during Early MAMP-Triggered Immunity. Plant Cell **29**: 20–38

Birkenbihl RP, Liu S, Somssich IE (2017b) Transcriptional events defining plant immune responses. Curr Opin Plant Biol **38**: 1–9

Bollina V, Kumarswamy GK, Kushalappa AC, Choo TM, Dion Y, Rioux S, Faubert D, Hamzehzarghan H (2010) Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley against Fusarium head blight. Mol Plant Pathol **11**: 76–82

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature **411**: 826–833

Dhokane D, Karre S, Kushalappa AC, McCartney C (2016) Integrated metabolo-transcriptomics reveals Fusarium head blight candidate resistance genes in wheat QTL-Fhb2. PLoS One. doi: 10.1371/journal.pone.0155851

Fry W (2008) Phytophthora infestans: The plant (and R gene) destroyer. Mol Plant Pathol **9**: 385–402

Fry WE (1978) Quantification of General Resistance of Potato Cultivars and Fungicide Effects for Integrated Control of Potato Late Blight. Phytopathology **68**: 1650–1655

Fry WE, Birch PRJ, Judelson HS, Grünwald NJ, Danies G, Everts KL, Gevens AJ, Gugino BK, Johnson DA, Johnson SB, et al (2015) Five Reasons to Consider *Phytophthora infestans* a Reemerging Pathogen. Phytopathology **105**: 966–981

Garner CM, Kim SH, Spears BJ, Gassmann W (2016) Express yourself: Transcriptional

regulation of plant innate immunity. Semin Cell Dev Biol **56**: 150–162

Gómez-Cano F, Soto J, Restrepo S, Bernal A, López-Kleine L, López CE (2019) Gene co-expression network for Xanthomonas-challenged cassava reveals key regulatory elements of immunity processes. Eur J Plant Pathol **153**: 1083–1104

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol **29**: 644–52

Gunnaiah R, Kushalappa AC, Duggavathi R, Fox S, Somers DJ, Young N, Danan S, Veyrieras J, Lefebvre V, Ballini E, et al (2012) Integrated Metabolo-Proteomic Approach to Decipher the Mechanisms by Which Wheat QTL (Fhb1) Contributes to Resistance against Fusarium graminearum. PLoS One **7**: 479–501

Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009) Applied biotechnology to combat late blight in potato caused by phytophthora infestans. Potato Res **52**: 249–264

Hussain RMF, Kim HK, Khurshid M, Akhtar MT, Linthorst HJM (2018a) Overexpression of AtWRKY50 is correlated with enhanced production of sinapic derivatives in Arabidopsis. Metabolomics **14**: 25

Hussain RMF, Sheikh AH, Haider I, Quareshy M, Linthorst HJM (2018b) Arabidopsis WRKY50 and TGA Transcription Factors Synergistically Activate Expression of PR1. Front Plant Sci **9**: 930

Jeger MJ, Viljanen-Rollinson SLH (2001) The use of the area under the disease-progress curve (AUDPC) to assess quantitative disease resistance in crop cultivars. TAG Theor Appl Genet **102**: 32–40

Jin J, Tian F, Yang DC, Meng YQ, Kong L, Luo J, Gao G (2017) PlantTFDB 4.0: Toward a central hub for transcription factors and regulatory interactions in plants. Nucleic Acids Res. doi: 10.1093/nar/gkw982

Jones JDG, Dangl JL (2006) The plant immune system. Nature **444**: 323–329

Kage U, Karre S, Kushalappa AC, McCartney C (2017) Identification and characterization of

a fusarium head blight resistance gene TaACT in wheat QTL-2DL. Plant Biotechnol J. doi: 10.1111/pbi.12641

Karre S, Kumar A, Dhokane D, Kushalappa AC (2017) Metabolo-transcriptome profiling of barley reveals induction of chitin elicitor receptor kinase gene (HvCERK1) conferring resistance against Fusarium graminearum. Plant Mol Biol. doi: 10.1007/s11103-016-0559-3

Katajamaa M, Orešič M, Ardrey R, Hoffmann E de, Stroobant V, Patterson S, Aebersold R, Goodacre R, Vaidyanathan S, Dunn W, et al (2005) Processing methods for differential analysis of LC/MS profile data. BMC Bioinformatics **6**: 179

Kushalappa AC, Gunnaiah R (2013) Metabolo-proteomics to discover plant biotic stress resistance genes. Trends Plant Sci **18**: 522–531

Kushalappa AC, Yogendra KN, Karre S (2016) Plant Innate Immune Response: Qualitative and Quantitative Resistance. CRC Crit Rev Plant Sci **35**: 38–55

Lewis LA, Polanski K, de Torres-Zabala M, Jayaraman S, Bowden L, Moore J, Penfold CA, Jenkins DJ, Hill C, Baxter L, et al (2015) Transcriptional Dynamics Driving MAMP-Triggered Immunity and Pathogen Effector-Mediated Immunosuppression in Arabidopsis Leaves Following Infection with Pseudomonas syringae pv tomato DC3000. Plant Cell **27**: 3038–64

Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics **12**: 323

Li B, Meng X, Shan L, He P (2016) Transcriptional Regulation of Pattern-Triggered Immunity in Plants. Cell Host Microbe **19**: 641–650

Liu S, Kracher B, Ziegler J, Birkenbihl RP, Somssich IE (2015) Negative regulation of ABA signaling by WRKY33 is critical for Arabidopsis immunity towards Botrytis cinerea 2100. Elife. doi: 10.7554/elife.07295

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and. Methods **25**: 402–408

Macho AP, Zipfel C (2014) Plant PRRs and the Activation of Innate Immune Signaling. Mol Cell

54: 263–272

Matías-Hernández L, Jiang W, Yang K, Tang K, Brodelius PE, Pelaz S (2017) AaMYB1 and its orthologue AtMYB61 affect terpene metabolism and trichome development in *Artemisia annua* and *Arabidopsis thaliana*. Plant J **90**: 520–534

Meraj TA, Fu J, Raza MA, Zhu C, Shen Q, Xu D, Wang Q (2020) Transcriptional Factors Regulate Plant Stress Responses Through Mediating Secondary Metabolism. Genes (Basel) **11**: 346

Musacchia F, Basu S, Petrosino G, Salvemini M, Sanges R (2015) Annocript: a flexible pipeline for the annotation of transcriptomes able to identify putative long noncoding RNAs. Bioinformatics **31**: 2199–2201

Nakano Y, Yamaguchi M, Endo H, Rejab NA, Ohtani M (2015) NAC-MYB-based transcriptional regulation of secondary cell wall biosynthesis in land plants. Front Plant Sci **6**: 288

Peng Y, Leung HCM, Yiu S-M, Lv M-J, Zhu X-G, Chin FYL (2013) IDBA-tran: a more robust de novo de Bruijn graph assembler for transcriptomes with uneven expression levels. Bioinformatics **29**: i326–i334

Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics **11**: 395

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci **14**: 21–29

Pushpa D, Yogendra KN, Gunnaiah R, Kushalappa AC, Murphy A (2013) Identification of Late Blight Resistance-Related Metabolites and Genes in Potato through Nontargeted Metabolomics. Plant Mol Biol Report **32**: 584–595

Ramšak Ž, Coll A, Stare T, Tzfadia O, Baebler Š, Van de Peer Y, Gruden K (2018) Network Modeling Unravels Mechanisms of Crosstalk between Ethylene and Salicylate Signaling in Potato. Plant Physiol **178**: 488–499

Shandil RK, Chakrabarti SK, Singh BP, Sharma S, Sundaresha S, Kaushik SK, Bhatt AK, Sharma NN (2017) Genotypic background of the recipient plant is crucial for conferring RB gene mediated late blight resistance in potato. BMC Genet **18**: 22

Sun T, Zhang Y, Li Y, Zhang Q, Ding Y, Zhang Y (2015) ChIP-seq reveals broad roles of SARD1 and CBP60g in regulating plant immunity. Nat Commun **6**: 10159

Taylor-Teeples M, Lin L, de Lucas M, Turco G, Toal TW, Gaudinier A, Young NF, Trabucco GM, Veling MT, Lamothe R, et al (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. Nature **517**: 571–575

Tsuda K, Somssich IE (2015) Transcriptional networks in plant immunity. New Phytol **206**: 932–947

Voiniciuc C, Yang B, Schmidt M, Günl M, Usadel B (2015) Starting to Gel: How Arabidopsis Seed Coat Epidermal Cells Produce Specialized Secondary Cell Walls. Int J Mol Sci **16**: 3452–3473

Wiesel L, Davis JL, Milne L, Redondo Fernandez V, Herold MB, Middlefell Williams J, Morris J, Hedley PE, Harrower B, Newton AC, et al (2015) A transcriptional reference map of defence hormone responses in potato. Sci Rep **5**: 15229

Wong DCJ, Ariani P, Castellarin S, Polverari A, Vandelle E (2018) Co-expression network analysis and cis-regulatory element enrichment determine putative functions and regulatory mechanisms of grapevine ATL E3 ubiquitin ligases. Sci Rep **8**: 3151

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al (2014) SOAPdenovo-Trans: de novo transcriptome assembly with short RNA-Seq reads. Bioinformatics **30**: 1660–1666

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, et al (2011) Genome sequence and analysis of the tuber crop potato. Nature **475**: 189–195

Yogendra KN, Dhokane D, Kushalappa AC, Sarmiento F, Rodriguez E, Mosquera T (2017) StWRKY8 transcription factor regulates benzylisoquinoline alkaloid pathway in potato conferring resistance to late blight. Plant Sci **256**: 208–216

Yogendra KN, Kumar A, Sarkar K, Li Y, Pushpa D, Mosa KA, Duggavathi R, Kushalappa AC (2015a) Transcription factor StWRKY1 regulates phenylpropanoid metabolites conferring late blight resistance in potato. J Exp Bot **66**: 7377–89

Yogendra KN, Kushalappa AC (2016) Integrated transcriptomics and metabolomics reveal induction of hierarchies of resistance genes in potato against late blight. Funct Plant Biol **15**: 497–7828

Yogendra KN, Kushalappa AC, Sarmiento F, Rodriguez E, Mosquera T (2015b) Metabolomics deciphers quantitative resistance mechanisms in diploid potato clones against late blight. Funct Plant Biol **42**: 284–298

Yogendra KN, Pushpa D, Mosa KA, Kushalappa AC, Murphy A, Mosquera T (2014) Quantitative resistance in potato leaves to late blight associated with induced hydroxycinnamic acid amides. Funct Integr Genomics **14**: 285–298

Zhang H, Yang B, Liu J, Guo D, Hou J, Chen S, Song B, Xie C (2017) Analysis of structural genes and key transcription factors related to anthocyanin biosynthesis in potato tubers. Sci Hortic (Amsterdam) **225**: 310–316

Zhang R, Zheng F, Wei S, Zhang S, Li G, Cao P, Zhao S, Zhang R, Zheng F, Wei S, et al (2019) Evolution of Disease Defense Genes and Their Regulators in Plants. Int J Mol Sci **20**: 335

Zhao W, Zhou X, Lei H, Fan J, Yang R, Li Z, Hu C, Li M, Zhao F, Wang S (2018) Transcriptional evidence for cross talk between JA and ET or SA during root-knot nematode invasion in tomato. Physiol Genomics **50**: 197–207

Zhou M, Memelink J (2016) Jasmonate-responsive transcription factors regulating plant secondary metabolism. Biotechnol Adv **34**: 441–449

Connecting statements to Chapter IV

Chapter IV is a manuscript titled **"Characterizing *StACT* genes and deciphering a signal cascade and its role in defense response in potato against *Phytophthora infestans*"**, which is authored by Sripad Joshi, Niranjan Hegde and Dr. Ajjamada C. Kushalappa.

Contributions of authors

Sripad Joshi conducted the experiments and wrote the manuscript. Niranjan Hegde helped with the tissue culture and pathogen inoculation. Ajjamada C. Kushalappa supervised the project and edited the manuscript.

In the previous study we tried to map the difference in the gene regulation between susceptible and resistant genotype. We obtained the metabolites, genes and transcription factors which play a role in conferring resistance to potato against *P. infestans*. Several metabolite classes were identified which were differentially accumulated between the two genotypes studied, however Hydroxycinnamic acid amides (HCAAs) are of more importance as previous studies in potato and other plant species it was identified that HCAAs play a crucial role in conferring resistance (Yogendra et al., 2015a; Kage et al., 2017b). Hence, we focused on this particular group, identified and characterized the specific metabolites, genes and transcription factors related to HCAAs. We tried to pinpoint a model related to specific metabolite and how tried to pinpoint the reason for its differential accumulation and find how the resistance changes after the mutated factor change can affect the resistance.

Chapter IV

Characterization of *StACT* genes and their regulation in potato to defend against *Phytophthora infestans*

Sripad Joshi[1], Niranjan Hegde[1] and Ajjamada C. Kushalappa[1]

1.Department of Plant Science, McGill University, Ste.-Anne-de-Bellevue, Quebec, Canada

4.1 Abstract

The hydroxycinnamic acid amides (HCAAs) are known to be induced following pathogen invasion and deposited to reinforce the secondary cell walls to contain the pathogen to an initial infection area. Metabolic profiling of resistant and susceptible genotypes, following pathogen inoculation, identified hydroxycinnamic acids and four specific types of amines namely, agmatines, cadaverines, putrescines and tyramines, however, only the feruloylagmatine was highly accumulated in the resistant genotype. The search for genes responsible for the accumulation of feruloylagmatine identified two agmatine coumaroyl transferase (ACT) genes, *StACT1* and *StACT2*. The sequence analysis of these genes revealed no mutations in the CDS regions of both resistant and susceptible genotypes. The *in-silico* protein docking studies revealed *StACT2* gene to be the functional isoform of the *ACT* that is responsible for the biosynthesis of feruloylagmatine in potato. Further search led to the identification of a transcription factor *StNAC72* that regulated *StACT2*, however, no mutation was revealed in the CDS of that gene. This led to the sequencing of promotor regions of *StACT2* and *StNAC72*, which revealed an 8 base pair deletion in the promoter region of *StNAC72*. This mutation was considered responsible for the reduced accumulation of feruloylagmatine in the susceptible cultivar, AG704. When this gene was edited, the gene expression significantly increased but not the disease severity. This was considered to be due to another missing link, in the hierarchy of regulatory and biosynthetic network of genes involved in the biosynthesis of the feruloylagmatine metabolite in the susceptible cultivar. Discovery and editing of another missing link should enhance feruloylagmatine, and thus the resistance to late blight.

4.2 Introduction

Late blight of potato is a devastating disease caused by the pathogen *Phytophthora infestans*. The quantitative resistance in plants is considered to be controlled by several genes. The disease response of a plant can be categorized into 3 distinct phases. It starts with the identification of the presence of an external pathogen by the plant cell receptors which are broadly classified as receptor proteins including the specific NBS-LRR proteins or can be broad spectrum like elicitors (Bent and Mackey, 2007). The second phase is where the disease response is mounted, and it involves several signaling mechanisms and transcription factors which play a major role in regulation (Tsuda and Somssich, 2015). In the third phase, a plethora of secondary metabolites are produced which have different mechanisms to suppress the invading pathogens (Kushalappa et al., 2016). These mechanisms may involve, antimicrobial activity or depositions to reinforce the cell walls in order to contain the pathogen to initial infection area (Kushalappa and Gunnaiah, 2013).

Hydroxycinnamic acid amides (HCAAs) is one class of secondary metabolites that have shown to work effectively to contain several pathogens in plants including Arabidopsis, barley, wheat, and others belonging to the (Macoy et al., 2015). Previous studies have shown significant effects of HCAAs against different pathogens, such as *Fusarium, Botrytis cinerea,* etc. (Kage et al., 2017b; Li et al., 2018). Specific compounds belonging to the HCAA class have more advantage over others to contain the pathogen. These compounds include the p-coumaroyl agmatine and feruloylagmatine. p-coumaroyl agmatine was initially identified in *Hordeum vulgare* and the hordatine was identified as its active form and it was initially found to be involved in cold stress (Burhenne et al., 2003). However, its antifungal properties were identified which had significant impact on the invading pathogen (Kage et al., 2017b; Carere et al., 2018). The structurally related compounds like feruloylagmatine are found to be functional in other plants like wheat (Jin and Yoshida, 2000). Feruloylagmatine and p-coumaroyl agmatine have shown antimicrobial effects in plant pathogen *in vitro* studies (Carere et al., 2018). They have also shown to be affecting the cell wall reinforcement which helps to contain the pathogen progress and in turn localizes the spread of the pathogen (Kage et al., 2017b).

Pathogen defense is a response mounted when the pathogen signal is perceived. Hence it is very important that the metabolites which are accumulated like the HCAAs are only activated after pathogen attack, as its constitutive accumulation might be unnecessary if there was no pathogen

attack. HCAAs belong to the phenylpropanoid pathway which conjugate with several other metabolites. These have the same precursor compounds and hence allocating these compounds can be tightly regulated by plants for the production of HCAAs. The specific agmatine amides have proven to be effective against pathogens like *Alternaria* in *Arabidopsis* however, no specific evidence of agmatine amides were studied in potato against *P. infestans* (Muroi et al., 2009). In large scale metabolomic experiments in potato- *P. infestans* interaction studies, several agmatine amines have been identified (Pushpa et al., 2013; Yogendra et al., 2015b). Hydroxycinnamoyl transferases form a set of large group of genes in the transferase family which catalyze the transfer of hydroxycinnamoyl group to the amine or amide group (Petersen, 2016). Hydroxycinnamoyl-CoA and the amines undergo a condensation reaction. Previously some tyramine hydroxycinnamoyl transferase was characterized in potato, however, no other agmatine related transferase was identified (Yogendra et al., 2015b). In the current study the conjugation of feruloyl-CoA and the agmatine amine which forms the feruloylagmatine conjugates that is known to be active in defending the pathogens was identified. Several other studies have shown that the agmatine coumaroyl transferase (*ACT*) gene also catalyzes the conjugation of feruloyl-CoA and agmatine along with coumaroyl-CoA and agmatine (Jin and Yoshida, 2000; Carere et al., 2018).

In this study HCAA conjugates are characterized which are active against the invading pathogen *P. infestans* in potato. The metabolic profile comparison of a late blight resistant and susceptible genotypes revealed the conjugated metabolites, which led to the search for the specific gene(s) which may be responsible for the accumulation of these conjugates. Two important genes that might be responsible for resistance were identified and designated as *StACT1* and *StACT2*. Along with these genes another gene that is Feruloyl transferase was also identified, which uses feruloyl-CoA as substrate. The gene expression, qPCR, and *in silico* protein docking identified possible involvement of *StACT2* gene in the feruloylagmatine accumulation in potato. This gene was found to be regulated by the *StNAC72* and the sequence analysis revealed a deletion in its promoter region in the genotype AG704. The editing of this gene in the susceptible AG704, however, failed to significantly enhance resistance. It is possible that other gene(s) involved in the network to biosynthesize feruloylagmatine may be mutated and editing of it may enhance the abundance of this metabolite in AG704, equal to that in resistant genotype Libertas.

4.3 Materials and methods

4.3.1 Identifying the hydroxycinnamic acid amides responsible for late blight resistance

From a previous metabolomics study, the conjugated amines were searched. The identified metabolites were plotted across 4 treatments; namely resistant genotype mock inoculated (RM), resistant pathogen inoculated (RP), susceptible mock inoculated (SM) and susceptible pathogen inoculated (SP) and the accumulation in each treatment was statistically compared with other using student's *t-test*. The potato resistant genotype used in this study was Libertas and the susceptible genotype used was an advanced breeding line AG704. The amines were filtered based on 4 important conjugates that is agmatines, cadavarines, putrescines and tyramines. The graphs were plotted where on the x-axis peak height was plotted, identified in the uHPLC (ultra-high-performance liquid chromatography) which was performed using LC-ESI-LTQ-Orbitrap-MS (ESI-Electron Spray Ionization, LTQ- Linear Trap Quadrapole, MS-Mass Spectrometry) (Thermo Fisher, USA) using a 5 cm kinetex column (Bollina et al., 2010). The metabolites were identified based on the observed m/z values that matched the database, with 5 ppm accurate mass error and were further confirmed using fragmentation patterns.

4.3.2 Identification of transferase genes

Two different approaches were used to retrieve transferase genes from potato which can be catalyzers of the hydroxycinnamic acid amides. Initially, fully characterized *Hordeum vulgare* agmatine coumaroyl transferase (*HvACT*) was used as query to identify homologous proteins from potato. The PHYTOZOME v8.0 database (www.phytozome.net/) and Ensembl Plants HHMER (https://plants.ensembl.org/hmmer/index.html) protein search tools were used with default parameters with *HvACT* as query against the potato genome as the filtered database. Also, the potato Spud DB (http://solanaceae.plantbiology.msu.edu/index.shtml) database was used to find the homologues sequences, using *HvACT* gene as the reference protein in potato genome.

4.3.3 Gene sequencing and identification of genes in potato genotypes

Genomic DNA was isolated from two genotypes that is Libertas and AG704 leaf samples following CTAB (Cetyltrimethylammonium Bromide, Molecular Biology Grade - CAS 57-09-0, from Millipore Sigma, Canada Inc.) DNA isolation protocol. 100mg young leaf tissue was subjected to fine grounding using liquid nitrogen in pre-chilled mortar and pestle. The ground tissue was then treated with CTAB buffer at 65^0C and DNA was isolated by phenol, chloroform method (Allen et al., 2006).

The PCR primers for 3 specific genes that is *StACT1*, *StACT2* and *StFT* were designed using primer BLAST at the NCBI server (Appendix VI). These primers were then used in the PCR reaction carried out using Q5 DNA polymerase (NEB Biosciences Inc. Canada) which also has a proof-reading activity. The amplified PCR product was sequenced using Sanger sequencing at the Genome Quebec sequencing facility at McGill University. The identified sequences were analysed using Snapgene Viewer and further translated into mRNA and protein sequences using the ExPasy translation tool (https://web.expasy.org/translate/). All the parameters were set to default and protein translation standard protein code was used.

For nomenclature, prefix 'St' for *Solanum tuberosum* was added to ACT, and FT genes and ACT genes were numbered according to close homology with *HvACT, as StACT1 and StACT2*. The chromosomal location of these genes was identified from potato database on potato chromosomes which was identified using the potato genome browser embedded with the PGSC database. To identify the domain and the family of the translated protein sequences Pfam database (http://pfam.sanger.ac.uk/) and the NCBI conserved domain database was searched using the protein sequences for all the 3 genes. The ScanProsite program from Expasy server (http://prosite.expasy.org/) was searched using the same protein sequences as query to identify specific domains and sites.

4.3.4 Quantitative reverse transcriptase PCR (qRT-PCR)

The total RNA was extracted from 4 treatments 48 hpi using a NucleoSpin RNA Plant kit (MACHEREY-NAGEL GmbH & Co., USA) according to the manufacturer's instructions. The obtained RNA was then converted to cDNA using the SMART® MMLV Reverse Transcriptase

(Takara Bio Inc., USA) using the polyT primers. Specific primers were designed for the 3 genes using the primer BLAST tool. The qRT-PCR reaction was carried out using these specific primers (Appendix IV) with 3 biological replicates and 3 technical replicates for each gene along with the housekeeping gene. Real time PCR reaction was carried out using IQ SYBR Green Supermix (BioRad, Hercules, CA, USA) with a CFX384 Real-Time System (BioRad, ON, Canada) according to the manufacturer's instructions. The relative gene expression level was calculated using the $2^{-\Delta\Delta Ct}$ method and Elongation factor 1-α (ef-1α) was used as the recommended housekeeping gene.

4.3.5 Phylogenetic tree construction

The amino acid sequences of *HvACT* was used as a query sequence in NCBI BLAST search, to identify homologues sequence in other plant species. The top hit from each species was selected and used to construct a phylogeny tree. Multiple sequence alignment (MSA) was performed by ClustalW2 and protein weight matrix BLOSSUM62 and default MSA Parameters. MEGAX software was used to plot the phylogeny trees. The evolutionary history of ACT genes was inferred by using the Neighbour Joining (NJ) method and 1000 bootstrap replicates were used for robust results.

4.3.6 Homology modelling of *StACT1* and *StACT2* proteins and substrate binding study

Protein homology-based structural modelling was performed using the translated protein sequences of the *StACT* genes. Homologues sequences were identified using the HHpred homology search against 3 databases namely, protein data bank (PDB), protein family (Pfam) and NCBI conserved domain (Zimmermann et al., 2017). Multiple sequence alignment was generated using the 12 homologues proteins from the HHpred search. The MODELLER software embedded with the MPI bioinformatics toolkit (https://toolkit.tuebingen.mpg.de/#/) was used to generate a protein model for both the genes (Webb and Sali, 2016). The protein models obtained were visualised using the PyMOL viewer (DeLano, 2002). The obtained protein structures were superimposed using the PyMol® tool to see the structural differences. The obtained structure was then used for docking the substrate molecules to find the differences in the substrate binding

affinities. Agmatine, feruloyl-CoA and coumaroyl-CoA molecular structure was downloaded from the ZINC database (http://zinc.docking.org/) and PubChem (https://pubchem.ncbi.nlm.nih.gov/) and the .mol2 files were used in docking.

AutoDockTools, which is distributed as part of the MGLtools package was used to dock the target protein and the substrate ligand molecules. The graphical interface program AutoDockTools-1.5.6 was used to set up the protein and the ligand to be docked, and the docking grid (Morris et al., 2009). The docking grid orientation was defined using the grid function around the binding site and the spacing (Angstrom) factor was set to 1.0 to have the measurements taken in Angstrom units. Multiple docking runs was done by running a batch AutoDock Vina via command line by using a script. The AutoDock Vina command line script was used were the grid box description and the names for the input and output files was mentioned (Trott and Olson, 2009).

4.3.7 Fluorescence assays to identify TF binding

To identify the NAC transcription factor binding to the ACT gene promoter, the *NAC72* cds sequence along with a 35S promoter was cloned which acts the effector protein and the native *StACT2* promoter along with a green fluorescent protein (GFP) gene was clones in a golden gate cloning vector with Esp3I cloning sites (Engler and Marillonnet, 2014). Along with the test vector another vector was similarly cloned which lacked the effector protein and only had the *StACT2* promoter and the GFP. A single tube reaction was designed to clone all the fragments in the test vector and the control vector with the Golden Gate cloning one tube reaction using the PCR thermo cycler. The protoplast from potato AG704.10 and Libertas young leaves was isolated following a protocol which was initially developed for *Arabidopsis* with some changes (Yoo et al., 2007; Yogendra et al., 2015a). The protoplasts were transformed with the test and the control vectors using the polyethylene glycol mediated transformation. The transformed protoplasts were viewed under the fluorescent microscope 48 post transformation for the GFP signal.

4.3.8 Genome editing using CRISPR Cas9 to insert the missing promoter element

CRISPR-Cas9 genome editing tool needs three important components (Jinek et al., 2012). Component A is the Cas9 protein driven by a specific promoter (CaMV 35s) and it works as the

nuclease enzyme to make the desired cut in the genomic DNA. The component B is the guide RNA which consists the CRISPR and tracer RNA specific to the region you desire to have a double strand break in the plant genomic DNA. The component C is the donor template with homology arms, and it works as the homology sequence during the homology directed repair which brings about the intended change in the native genomic DNA. To construct the desired plasmids, we followed golden gate cloning toolkit which is customizable according to the plant and single or multiple gRNA (Čermák et al., 2017). The components A, B and C were cloned in a final destination T-DNA vector which was used in the *Agrobacterium* mediated plant transformation. The component A (pMOD_A0501) had the *AtCas9* driven by a CaMV35 promoter. We used a multiple guide RNA system which are separated by Cys4 systems and the vector used was pMOD_B2102. The gRNAs were designed using the CRISPR-P 2.0 (http://crispr.hzau.edu.cn/CRISPR2/) and selecting the off target screening using the potato database (Appendix VI). Guide RNA with high scoring and no other off targets were selected as gRNA. The donor DNA was designed such that 500bp on the flanking ends of the gRNA binding was selected for homology directed repair and Kanamycin resistance gene nptII along with 35S promoter was cloned in the C vector. The 3 components A, B and C were cloned in the final destination T-DNA vector pTrans_201 which also had the Gemini viral replicons.

4.3.9 *Agrobacterium* mediated transformation and selecting the positive transformants

The potato internodes were used as explant for *Agrobacterium* mediated transformation. *Agrobacterium* GV3101 strain was transformed with the final pTrans_201 vector using the heat shock method. Following transformation, the internodes were cocultivated with the *Agrobacterium* containing the transformed vector and further inoculated on the callus induction medium (MS medium supplemented with 3% sucrose, 3 mg/L benzylaminopurine, 2 mg/L naphthalene acetic acid, and 6 g/L agar) supplemented with antibiotic Augmentin. After 2 weeks antibiotic Kanamycin was also added to select the positive transformants. The plates were placed in the growth incubator at 25 °C with 16 hours of light. The plants were regenerated from the callus using the shoot induction medium (MS medium supplemented with 3% sucrose, 2.5 mg/trans-zeatin, 0.3 mg/L gibberellic acid GA3, 6 g/L agar, containing Augmentin and Kanamycin) and the regenerated plants were planted in soil 8 weeks after inoculating in shoot induction medium (De

Block, 1988). To detect the positive transformants primers were designed, differentiating the desired change (Appendix VI). The PCR product were further sequenced to confirm the desired change. The positive transformants were used to multiply plants and test for disease resistance.

4.3.10 Disease severity analysis

The detach leaf disease severity assay was done for the control AG704 and Trans AG704 leaves in the lab. The *P. infestans* sporangial concentration in the suspension was adjusted to $1X10^5$ spores per ml and 10ul spore suspension was inoculated using the Hamilton syringe on either side of the midrib on the leaves. For control samples mock inoculations were done using 10ul of sterile water. The experimental units consisted of three pots, with one plant in each, for each treatment. The leaves were picked randomly from all 6 control and transgenic plants.

4.4 Results

4.4.1 HCCAs were identified as the defense metabolites

The comparison of metabolic profiling data of two potato genotypes, a late blight resistant (Libertas) and another susceptible (AG704), inoculated with mock or *P. infestans* identified metabolites from several pathways. However, for the present study only the specific HCAA which are known be responsible for quantitative resistance were focused. Specific conjugates of HCAAs, such as, agmatines, cadaverines, putrescines and tyramines were searched. The abundances of these metabolites, the peak heights, were statistically compared in 4 treatments (Fig. 4.1). Two important conjugates of agmatine, feruloylagmatine and p-coumaroylagmatines were identified. The feruloylagmatine accumulated in significantly higher abundances in RP than in SP (Fig 4.1 A). However, the differences in p-coumaroyl agmatine between RP and SP were not significant.

None of the cadaverine conjugates were significant in any of the treatments. Other important conjugates identified were: N-feruloyltyramine, methoxytyramine-betaxanthin, feruloylputrescine and N-caffeoylputrescine. Although the difference in accumulation of these metabolites between RP-RM and SP-SM treatments was significant ($p<0.05$) it was not found to be significant between resistant and susceptible pathogen inoculated treatments. Hence, it was concluded that no other

HCAAs other than feruloylagmatine may be responsible for fostering resistance in Libertas genotype.

4.4.2 Genes regulating HCAA accumulation

Transferases gene family is a large group of genes found in bacteria, plants and animals, which can transfer a large moiety to another compound. It is evident that HCAA's are catalyzed by transferase genes which conjugates an acyl donor molecule to an acyl acceptor. To identify transferases in potato, a keyword search with "transferase" as a query and found 231 entries in SPUD db. However, there are several classes of transferase genes and all of them do not catalyze HCAA formation. Further, the search was restricted only pertaining to the catalysis of feruloylagmatine.

The homologues of *HvACT* in potato were searched in three different databases, SpudDB (http://solanaceae.plantbiology.msu.edu/), Phytozome (https://phytozome.jgi.doe.gov/pz/portal.html), and EnsemblPlants (https://plants.ensembl.org/index.html). The results were consistent with all 3 databases and 2 highest similarity proteins were found in potato. The highest similarity was with PGSC0003DMP400032595 and PGSC0003DMP400012693 with top query cover and 52.15 and 52.13 percent identity respectively (Table 4.1). Further investigation identified that these two genes were annotated as Anthranilate N-benzoyltransferase protein in potato database. The identified proteins have not been characterised in potato and only putatively characterised and annotated as anthranilate N-benzoyltransferase, however, the protein homology search identified them as closest to the agmatine coumaroyl transferase.

4.4.3 Characterization of agmatine coumaroyl transferase *(ACT)* genes in potato

The putatively characterised *ACT* genes in potato were sequenced in genotypes AG704 and Libertas. In addition, the feruloyl transferase gene was also sequenced because while searching for the transferases, where acyl donor is feruloyl, this gene was also detected (PGSC0003DMP400054926).

The genomic DNA was isolated from leaf tissue of AG704 and Libertas. The primers were designed using Primer BLAST and PCR was performed using primers for 3 different genes (Appendix VI). The sequences obtained were analysed using Snapgene Viewer and the chromatograms were analysed using ApE sequence analyser to find any aberrations in the obtained sequence. The *StACT1* and *StACT2* genes showed no intronic sequences and it was a complete single exon. The size of the potato *StACT1* and *StACT2* genes as shown in the SpudDB was 1323 base pairs with single transcript having no introns (Fig 4.2).

However, the sequence results showed some differences in the sequence with respect to AG704 and Libertas. In AG704, *StACT1* and *StACT2* was found to be 1323 base pair. *StACT1* in Libertas was 1323 and *StACT2* was 1326 base pairs. There was a 3 base pair difference in between AG704 and Libertas *StACT2* which should correspond to a single amino acid in protein (Fig 4.2 C). To further analyse the sequence, the DNA sequence was translated into protein coding using a standard eukaryotic codon table. The obtained protein sequence was aligned to each other to find if there are any major sequence aberrations at protein level (Fig 4.2 B & D).

StACT1 showed different amino acids at positions 38, 66, 83, 222 and 388 (Fig 4.2 A). Whereas, *StACT2* showed an additional Aspargine residue at position 255 (Fig 4.2 C). There was amino acid difference found at positions 27, 251, 280 and 346 (Fig 4.2 C). The differences did not correspond to a major difference in the amino acid class especially in *StACT1* and *StACT2* (Fig 4.2 B & D). The CDS sequence of the *StFT* did not show any difference in amino acid even though a few single nucleotide polymorphisms were identified. Although these changes should not affect the protein sequence and structure.

4.4.4 Domain identification and protein folding

To further analyze the amino acid differences in *ACT* genes, a motif and domain search was performed. There were no particular motifs identified in both the *ACT* genes, however, both the genes were grouped into plant transferase family, specifically to the PLN02481 domain super family that comprises the omega-hydroxy palmitate O-feruloyl transferase (Fig 4.3C). Hence, there was no difference found in the domains in both the genotype due to the amino acid differences.

The protein difference was further analysed using the 3-dimensional protein structure. The *StACT1* and *StACT2* protein sequences from both the genotypes were used to find the protein 3D structure by homology modelling. The AG704 and Libertas *StACT1* and *StACT2* did not show any structural aberrations due the amino acid changes (Fig 4.2 A & B). To compare the sequences of *StACT1* and *StACT2,* both the sequences were aligned to each other to see the level of differences between these genes (Fig 4.3 A). Several amino acids were found different between *ACT* genes but when the 3D protein structure shows high level of similarity (Fig 4.3 B). Structural similarity showed that both the genes should be similar in their functionality.

4.4.5 Quantifying the gene expression

The amino acid sequence of the *StACT1* and *StACT2* protein did not show any difference and also the structural differences did not show any significant change. To further investigate how these genes are expressed and to understand their function in greater details, a quantitative reverse transcriptase PCR (qRT-PCR) experiment was performed using unique primers designed for each gene (Appendix VI). The total RNA isolated from 4 treatments, with 3 biological replicates at 48 hpi and 3 technical replicates were used. The qPCR data showed *StACT1* and *StFT* had no significant difference between treatments, in both the genotypes (Fig 4.4 A & B). However, the *StACT2* showed significantly higher expression in resistant relative to susceptible genotype, following pathogen inoculation (FC = 2.25) (Fig 4.4 B). The difference was significant between RP and SP and also between RP and RM (FC=4.2), which showed that this particular gene is only induced after pathogen attack. However, the same induction was not significant in the susceptible genotype, which correlated with the difference in feruloylagmatine accumulation.

4.4.6 Phylogenetic tree construction

To better understand the relation of potato transferase genes with other homologues genes in plants a phylogenetic analysis was constructed. The *HvACT* sequence was used as a query to search the NCBI protein database and the phytozome protein BLAST to extract the similar sequences in other plants. The *HvACT* gene instead of potato was used because there was only 54% identity in *HvACT* and *StACT* protein sequences, but the intent is to identify the relativeness to most plant species.

Hence the BLAST search results were further filtered to identify the closest homologue in each species to have an expanded view of the phylogeny in the plant kingdom. The identified protein sequences from plants were further grouped according to their order in the Angiosperm phylogenetic group (APG) 3 classification system. The multiple sequence analysis was performed with the two *StACT* and *StFT* genes and phylogenetic tree was constructed using MEGA X.

The *StACT1* and *StACT2* grouped in a single clade with the other Solanales members however, they were distinctly divided into two separate subgroups (Fig 4.5). This showed that the *ACT* genes are similar, yet they have certain differences between them when it comes to the evolutionarily closed species. The other order species showed some specific clades including the Poales, Malphigiales and Aerales (Fig 4.5). However, the other species did not show any specific clades, which showed that there was no distinct evolutionary difference that can be correlated with respect to the *ACT* genes. It was interesting to note that Brassicales, Vitales, Fabales and Cucurbitales were grouped together and did not form separate clades.

4.4.7 *In silico* protein-substrate docking

The *ACT* genes act as an enzyme catalyzing a specific reaction with its substrates. The substrates have to specifically bind to the protein to facilitate catalysis. In this reaction the agmatine is the amine substrate and along with the agmatine the other important substrate is the hydroxycinnamoyl-CoA. Accordingly, the binding affinities of feruloyl-CoA, coumaroyl-CoA and agmatine were explored with the homology modelled protein structures of *StACT1* and *StACT2*.

The agmatine binds to the inner pocket of the folded protein and the docking studies showed some difference in the binding affinity with small difference in ΔG. The binding affinity was recorded as -4.4 and -5.1 kcal/mol in *StACT1* and *StACT2* respectively, however, the binding sites were identical, which is deep inside the pocket (Table 4.2). The feruloyl-CoA and coumaroyl-CoA showed significant difference in the binding affinity and as well the binding site was different between *StACT1* and *StACT2*. In case of coumaroyl-CoA, the binding affinity of the lowest free energy binding site was -8.0 and -9.3 kcal/mol for *StACT1* and *StACT2,* respectively (Table 4.2). Whereas for feruloyl-CoA in *StACT1* it was found -6.9 kcal/mol and for *StACT2* it was found -9.0

kcal/mol (Table 4.2). The difference in the binding affinity shows that there is stark difference in the binding affinity of the hydroxycinnamoyl-CoA. When the binding between the substrates and the folded protein was visualized using the PyMOL viewer, the binding sites were different for both the hydroxycinnamoyl-CoA between *StACT1* and *StACT2* (Fig 4.6). In case of *StACT2* the feruloyl-CoA and coumaroyl-CoA ligand is seen to be docked exactly inside the binding pocket which has a higher binding affinity (Fig 4.6 A and B). Whereas the binding of feruloyl-CoA and coumaroyl-CoA ligands to *StACT1* showed the lowest binding was not in the binding pockets but was in the shallow (Fig 4.6 C and D) region on the surface of the protein. It was also evident to note that agmatine binds inside the pocket in both the genes. Hence the other substrate has to bind in the same binding pocket to facilitate the catalytic reaction. Therefore, the difference in the binding shows that *StACT2* is the functional gene which can successfully carry the transferase enzymatic activity, whereas *StACT1* is the non-functional isoform.

4.4.8 Promoter analysis and transcription factor binding assay

The promoter of *StACT2* in AG704 and Libertas was sequenced, but no difference in the sequence was recognized except a few single nucleotide changes. When the 1500bp upstream promoter sequence of *StACT2* gene was analyzed, several transcription factors binding sites were found. But when these TFs were correlated with the previous RNAseq data, only *StNAC72* was found differentially expressed in Libertas and AG704. The *StNAC72* binding site was found at -1164bp (CGTAGATTGAAATGAAAGTAA) upstream of start codon (ATG). The *StNAC72* gene expression was confirmed using the qPCR in 4 treatments and was found differentially accumulated in the resistant pathogen inoculated (RP) treatment (FC=6.3), which further gave an understanding that this gene might be responsible for the differential regulation of *StACT2* (Fig 4.7 A).

To experimentally confirm the binding of *StNAC72* to *StACT2* promoter, a binding assay was designed where the *StNAC72* was driven by a CaMV 35S promoter and *StACT2* promoter cloned with a GFP (Fig 4.8 A). The GFP expression was not seen in the control (Fig 4.8 D) samples which only had *StACT2* promoter with GFP (Fig 4.8 C) in the fluorescence microscopy. The fluorescence signal was clearly seen in the protoplasts transformed with the test vectors which had the *StNAC72* as the effector along with the GFP with *StACT2* promoter (Fig 4.8 B). This confirmed the binding

of *StNAC72* to the promoter of *StACT2* which was seen as fluorescence signal in test protoplasts and not in the control.

When *StNAC72* gene was sequenced no difference in the sequence was found among genotypes used here, and the domain region did not show any difference at all. However, an 8-base pair (CGTGTCAA) deletion was detected in the promoter sequence of the *StNAC72* in AG704*,* which was 835bp upstream of the start codon. This mutation which could be the reason for downregulation of *StNAC72* and in turn *StACT2* in AG704 than Libertas post pathogen inoculation. Based on this a regulation model of *StACT2* (Fig 4.7 B) was designed, where *StNAC72* regulates the expression of *StACT2* in potato. The mutation in the promoter region of *StNAC72* (Fig 4.7 B), corresponds to the bZIP transcription binding site. Seven possible bZIP TFs were identified with binding site analysis as possible targets. However, when these were cross referenced with the RNAseq data none of them showed differential expression. Hence, it was evident that the mutation in the promoter of *StNAC72* was the limiting factor in AG704, and it led to the differential gene expression.

4.4.9 Generating transgenic AG704 plants with edited mutation

The mutation found in the promoter of the *StNAC72* gene was further explored to how it can affect the gene expression and resistance in general. A genome editing experiment was designed wherein the missing bases in AG704 could be replaced based on Libertas sequence. To facilitate easy selection of positively edited plants, the Kanamycin resistance gene cassette was inserted as a selectable positive selection marker (Fig 4.8 B) 200 bp upstream of the site of mutation. Confirmation primers (Appendix VI) were designed outside the inserted cassette which has the Kanamycin gene and the homology arms, to select only the correct transformed plants. Plants showing single band PCR product with inserted cassette were selected as positive transformants which should be the homozygous plants with respect to the edited mutation and the plants showing lower size bands or double bands were eliminated as they should be heterozygous. In a total of 32 plants selected as transformed positives, only 6 had correctly inserted donor DNA which did not show any other band other than the desired band and these were further analyzed and sequenced to confirm the desired insertion (Fig 4.8 A).

4.4.10 Quantifying the disease severity and gene expression in the transgenic and the control plants

The transgenic plants obtained were further grown in soil inside the growth chambers along with the control plants. The experiment was designed as a randomized complete design with detached leaves collected randomly from all transgenic and control plants and were placed in Petri dish lined with moist filter papers and inoculated with either *P. infestans* spores or water as mock, in three replicates. Hence, there were 4 treatments in total including the TransAG704 mock and pathogen inoculated and ControlAG704 with mock and pathogen inoculated. The lesion diameter was measured every three days, for 9 dpi, using Vernier caliper and the average lesion diameter was 36.7mm in TransAG704 and 42.2 mm in the control AG704 plants (Fig 8A) at 6dpi, but with time there was not much difference. The AUDPC was calculated 120.6 in TransAG704 and 136.4 in the control plants. However, the difference in the AUDPC was not significantly different in TransAG704 from the control plants (Fig 4.9 B).

To find if editing the mutated segment had any effect on the gene expression, the total RNA was extracted from 4 treatments at 48 hpi, that is transgenic and control plants inoculated with mock and pathogen. qPCR was done with the same primers previously used to find the gene expression for *StACT2* and *StNAC72* (Appendix VI). For all the treatments 3 biological replicates and 3 technical replicates were used. Both the *StNAC72* and *StACT2* genes had higher expression in TransAG704 than in the control plants post pathogen inoculation and there was no difference seen in the controlAG704 plants (Fig 4.9 C).

4.5 Discussion

4.5.1 Feruloyl agmatine acts as an active HCAA in late blight resistance

The metabolites play a significant role in growth, development and structural makeup of plants. The phenylpropanoid pathway is central to the secondary metabolites and is vital for the innate immunity of plants against invading pathogens along with other pathways. The resistance in plants against pathogen attack is mainly due to resistance metabolites that are biosynthesized in plant metabolic pathway network. They are not only antimicrobial but also the polymers and conjugated forms are deposited to reinforce the secondary cell walls around the pathogen infected area.

HCAAs are further formed mainly by conjugation of end products of these secondary metabolite pathways. The feruloylagmatine is a conjugated metabolite (HCAA) of a hydroxycinnamic acid and agmatine as amine, which is biosynthesized from amino acids arginine. In this study feruloylagmatine, was found in higher abundances (FC=2.34) in a resistant genotype Libertas than in a susceptible AG704 after pathogen inoculation. Previous studies in potato have found other HCAA like feruloyl putrescine, N-feruloyl tyramine, N-caffeoyl putrescine, 4-coumaroylagmatine, p-coumaroyl tyramine, and others, along with feruloyl agmatine (Yogendra et al., 2014; Yogendra et al., 2015a). Also, in other plants these metabolites were found to be accumulated post pathogen inoculation (Muroi et al., 2009; Kage et al., 2017b; Carere et al., 2018).

The HCAAs are the conjugated metabolites biosynthesized by the transferases, ligases and other classes of enzymes. Among these the terminal transferase is called as the hydroxycinnamoyl transferases. Several hydroxycinnamoyl transferase have been characterized in *Arabidopsis,* wheat, tobacco, and barely (Muroi et al., 2009; Onkokesung et al., 2012; Kage et al., 2017b; Peng et al., 2019). In potato several HCAA metabolites were identified following pathogen invasion, however, none of the hydroxycinnamoyl transferase were previously characterized with respect to the agmatine amides (Yogendra et al., 2015b). The transferase family of genes which catalyzes the formation of HCAAs are named after the first four enzymes detected from this family, namely benzyl alcohol O-acetyltransferase, anthocyanin O-hydroxycinnamoyl transferase, anthranilate N-hydroxycinnamoyl/benzoyltransferase and deacetylvindoline 4-O-acetyltransferase (Petersen, 2016).

4.5.2 *StACT2* is the functional isoform of *ACT* in potato

Agmatine containing HCAAs were found in *Hordeum vulgare* in which coumaroyl agmatine was the first characterized and also an important gene *ACT* was identified which catalyzes the reaction using the coumaroyl-CoA and agmatine (Burhenne et al., 2003). In case of *ACT,* it catalyzes the reaction where agmatine which acts as acyl donor is transferred on to the coumaroyl group which is the acyl acceptor, with the release of CoA and a proton (Burhenne et al., 2003). A few other studies also reported that the same *ACT* enzyme also catalyzes feruloyl and caffeoyl moiety substrates as acyl acceptors. Hence, the *HvACT* which is a well characterised gene was used in our query search protein to identify homologues in potato. Two important transferase genes were characterized in this study which could catalyze the formation of feruloylagmatine that is *StACT1*

and *StACT2*. They were initially annotated as anthranilate N-hydroxycinnamoyl/ benzoyltransferase in the SPUD database. However, its close relation to the *HvACT* gene suggests that it should be reannotated and specific nomenclature should be used for these transferases. qPCR expression showed that *StACT2* is upregulated post pathogen inoculation in resistant genotype and not in susceptible. Further the *in-silico* substrate docking showed that *StACT2* is the functional form and not *StACT1*. Similarly in *Brachipodium, ACT* gene was found catalyzing reactions involving agmatine, putrescine, and cadaverine amines and shows broad spectrum specificity with hydroxycinnamic acid during *Fusarium* infection (Carere et al., 2018). Also, in wheat against *Fusarium* infection *ACT* was found functional and its silencing lead to increase in susceptibility (Kage et al., 2017a).

4.5.3 *StNAC72* regulates the expression of *StACT2* post pathogen infection

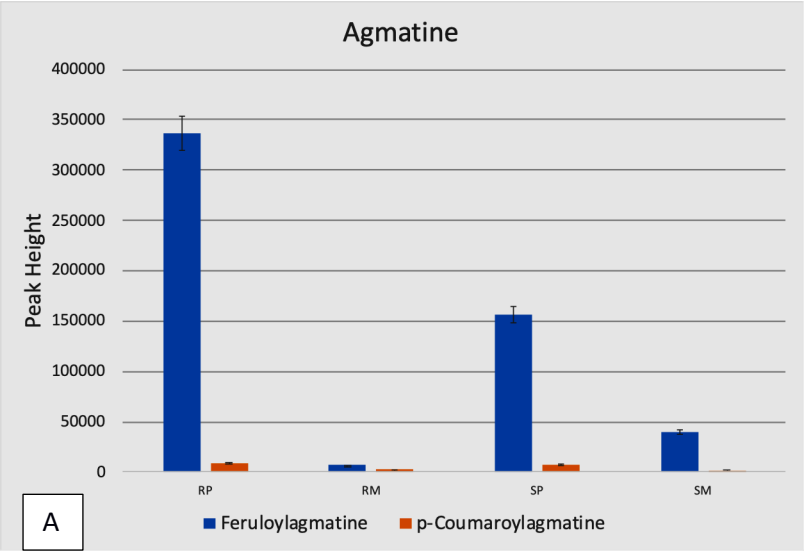
The *ACT* genes are known to be regulated by several TFs including *MYB, WRKY, ORA* etc. (Onkokesung et al., 2012; Kage et al., 2017b; Li et al., 2018). In this study, the TF *StNAC72* was identified as one of the regulators of the *StACT2* based on gene expression and its abundance was higher (FC=6.3) in resistant than in susceptible genotype, post pathogen inoculation. The sequence revealed no mutation in the resistant genotype but in the susceptible, though the exon region was intact, there was a mutation in the promotor region, a deletion of 8 base pairs. An attempt was made to edit the mutation in susceptible genotype using the sophisticated genome editing technique, CRISPR Cas9. It was found that the TransAG704 which has the edited promoter sequence of *StNAC72,* showed increased expression of both the genes, *StNAC72* and *StACT2* compared to the control plants. However, the disease severity showed no significant difference. It is possible that other genes involved in the pathway which leads to resistance could be mutated and those genes may be involved in lower resistance in AG704. The gene regulation model presented here would allow exploration of other TFs that may regulate *StACT2*.

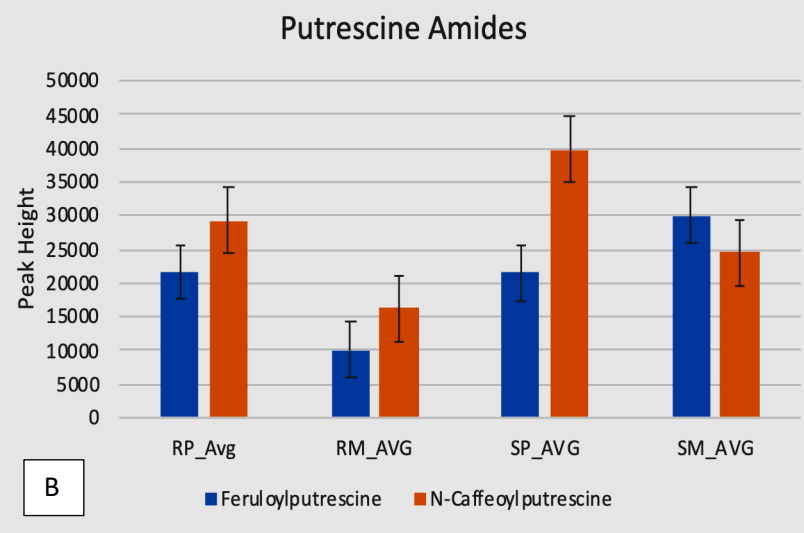
The homologues TFs of *StNAC72* in tomato are involved in response to *Botrytis cinerea* and is directly related to the jasmonic acid pathway (Du et al., 2017) and it was noted in other species like *Arabidopsis* that *ACT* genes are regulated by the jasmonic acid pathway (Li et al., 2018). It was identified that ORA59 regulates the *ACT* and it was directly linked to the jasmonic acid synthesis during stress response (Li et al., 2018). Hence, further studies which can incorporate the

jasmonic acid quantification in the potato genotypes and linking it to the *StNAC72*, *StACT* and other related genes can decipher several more possibilities to increase the resistance many folds. It is also known in potato that several *NAC* gene family members play an antagonistic role in suberin production, however, they in turn increase metabolites like feruloyl esters and alkanes. This shows that *NAC* transcription factors play a crucial role in the biosynthesis of phenyl propanoids and other secondary metabolites, either promoting or suppressing certain metabolite (Soler et al., 2020). Several new forms of HCAAs are also been identified in plants which will add more knowledge in understanding the role of these compounds in plant defense (Voynikov et al., 2016). Similarly, new studies should be undertaken to identify the genes that are responsible in catalyzing these metabolites. The metabolite biosynthetic genes may be regulated by primary TFs and these are also regulated by secondary TFs, which in turn regulate the downstream genes (Chapter III). Thus, the other TFs that may also regulate *StACT2*, and also probably the *StNAC72*, but may be mutated, should also be explored.

4.6 Conclusion

The feruloylagmatine is an important metabolite which has shown to be important for plant defense and other growth and development functions. *ACT* genes also play a crucial role in the defense by stimulating the formation of these important metabolites and it was important to identify the functional *ACT* gene in potato which was never explored. Hence, this study identifies that *StACT2* is the functional gene which can be further explored to know how it plays a role in different genotypes varying in disease resistance. A gene regulation model was identified where *StNAC72* regulates the *StACT2* gene post pathogen infection. A mutation in the *StNAC72* promoter region in the susceptible genotype was edited to increase the *StNAC72* and *StACT2* gene expression, however, the higher gene expression did not increase the resistance significantly. The other reason could be that the other transcription factor or microRNA in the hierarchy of feruloylagmatine could be responsible for resistance and its identification and editing could increase the resistance.


Agmatine
400000
350000
300000
250000
200000
150000
100000
50000
0
Peak Height
RP
RM
SP
SM
A
Feruloylagmatine
p-Coumaroylagmatine


Putrescine Amides
50000
45000
40000
35000
30000
25000
20000
15000
10000
5000
0
Peak Height
RP_Avg
RM_AVG
SP_AVG
SM_AVG
B
Feruloylputrescine
N-Caffeoylputrescine

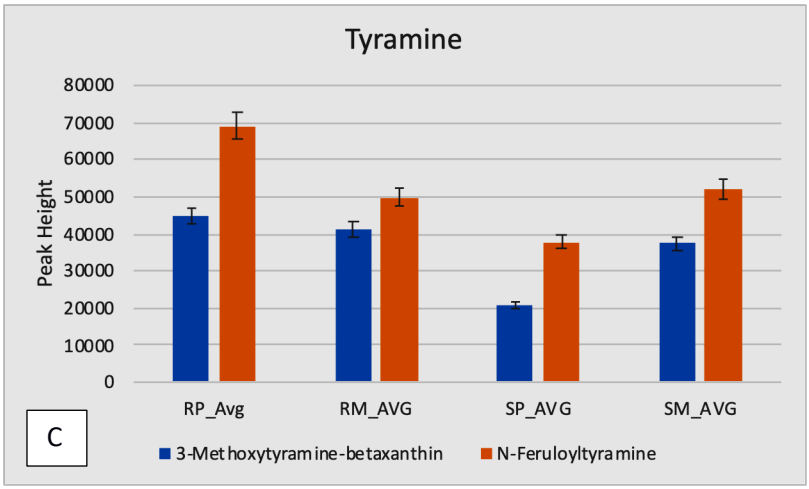

Tyramine
80000
70000
60000
50000
40000
30000
20000
10000
0
Peak Height
RP_Avg
RM_AVG
SP_AVG
SM_AVG
C
3-Methoxytyramine-betaxanthin
N-Feruloyltyramine

Fig 4.1 HCAA conjugates accumulated in RP, RM, SP and SM.

Hydroxycinnamic acid amides (HCAAs) in late blight resistant (Libertas) and susceptible (AG704) potato genotypes, following mock and pathogen inoculation, identified based on metabolic profiling: A) The differential accumulation of feruloyl agmatine and p-coumaroylagmatine in 4 treatments that is RP, RM, SP, SM. The feruloylagmatine is accumulated at a higher level in RP compared to other treatments where p-coumaroylagmatine accumulation shows no difference in accumulation in all treatments. B) Shows the accumulation of feruloylputrescine and N-caffeoylputrescine which show no difference in the level of their accumulation in the treatments. C) Shows no significant difference in the accumulation of 3-methoxytyramine-betaxanthin and N-feruloyltyramine in all 4 treatments. R is resistant, S is susceptible, P is pathogen and M is mock inoculated.

Table 4.1 Homologues genes to *HvACT* in potato.

Genomic Location	Overlapping Gene(s)		Number of Exons	Protein Length	Length	Score	E-val
11:39493132-39493734	PGSC0003DMG400000429	PGSC0003DMT400001145	1	441aa	201 [Sequence]	612	3.90E-149
6:55234818-55235111	PGSC0003DMG400007155	PGSC0003DMT400018441	1	445aa	98 [Sequence]	322	2.60E-138
11:42878971-42879231	PGSC0003DMG400018699	PGSC0003DMT400048126	3	451aa	87 [Sequence]	306	1.20E-127
11:42907443-42907703	PGSC0003DMG400018700	PGSC0003DMT400048131	1	443aa	87 [Sequence]	305	1.50E-126
		PGSC0003DMT400048130	2	569aa			
3:57307487-57307765	PGSC0003DMG400014152	PGSC0003DMT400036695	2	426aa	93 [Sequence]	203	1.50E-34
7:1002296-1002574	PGSC0003DMG400011189	PGSC0003DMT400029079	2	360aa	93 [Sequence]	200	2.20E-30
		PGSC0003DMT400029078	1	203aa			
3:46127922-46128203	PGSC0003DMG400031731	PGSC0003DMT400081182	2	433aa	94 [Sequence]	181	4.50E-21
		PGSC0003DMT400081183	1	206aa			

The homologue genes that are identified in potato with the query sequence input that is *HvACT* in EnsemblPlants BLAST program. We can find less identity in the top hit sequences; however, the scores are high, which says that the amino acid sequence is not identical its largely similar.

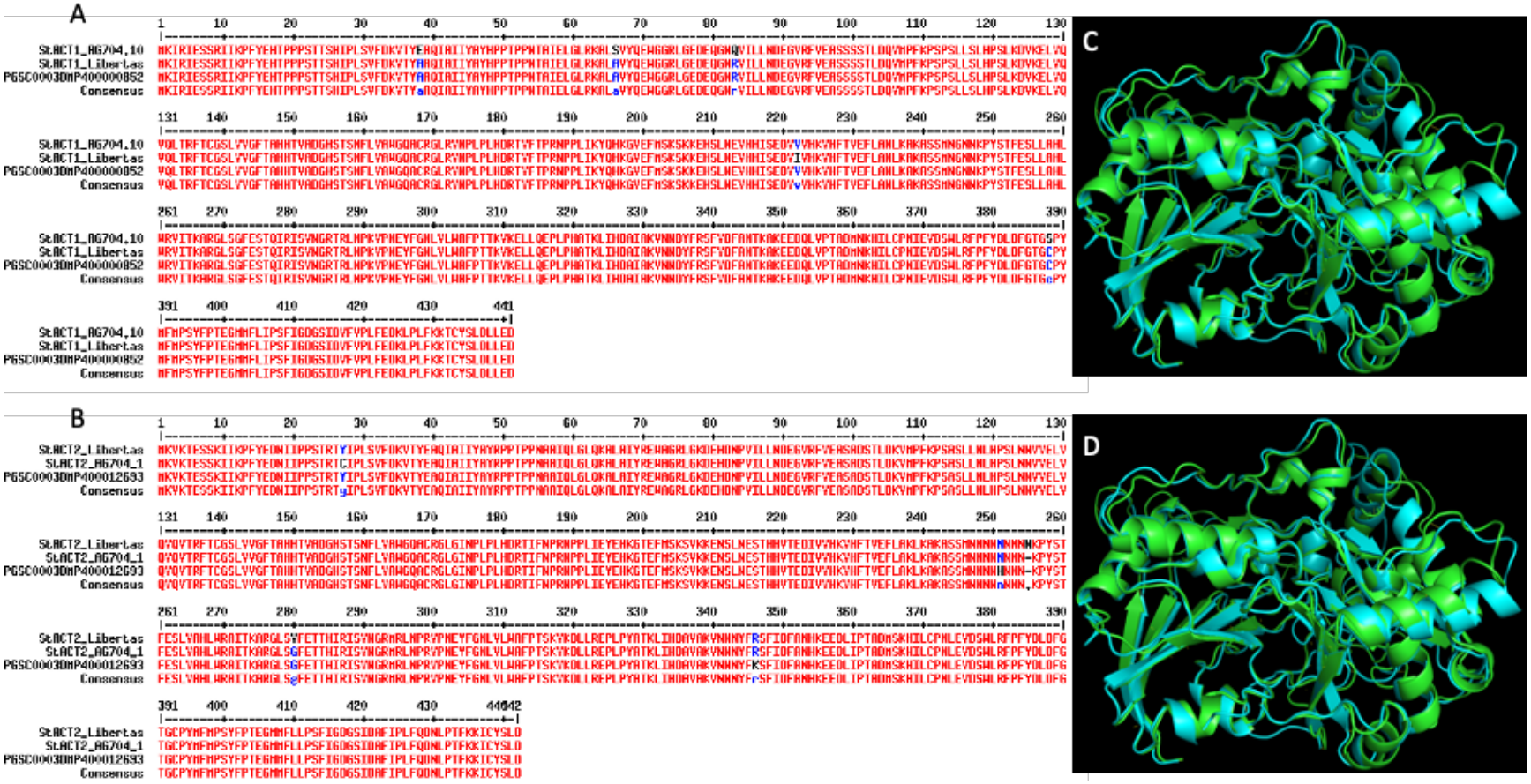


Fig 4.2 *StACT1* and *StACT2* protein comparison between AG704 and Libertas.

The translated protein sequence between the genotypes AG704 and Libertas. A and B show the *StACT1* and *StACT2* protein sequences respectively, where it can be seen that some amino acid changes at position 38, 66, 83, 222 and 388 in *StACT1*; and positions 27, 251, 280 and 346 in *StACT2.* In the figures C and D, the homology modelled protein structures are shown for *StACT1* and *StACT2* and the two genotypes protein folding structures are superimposed for comparison.

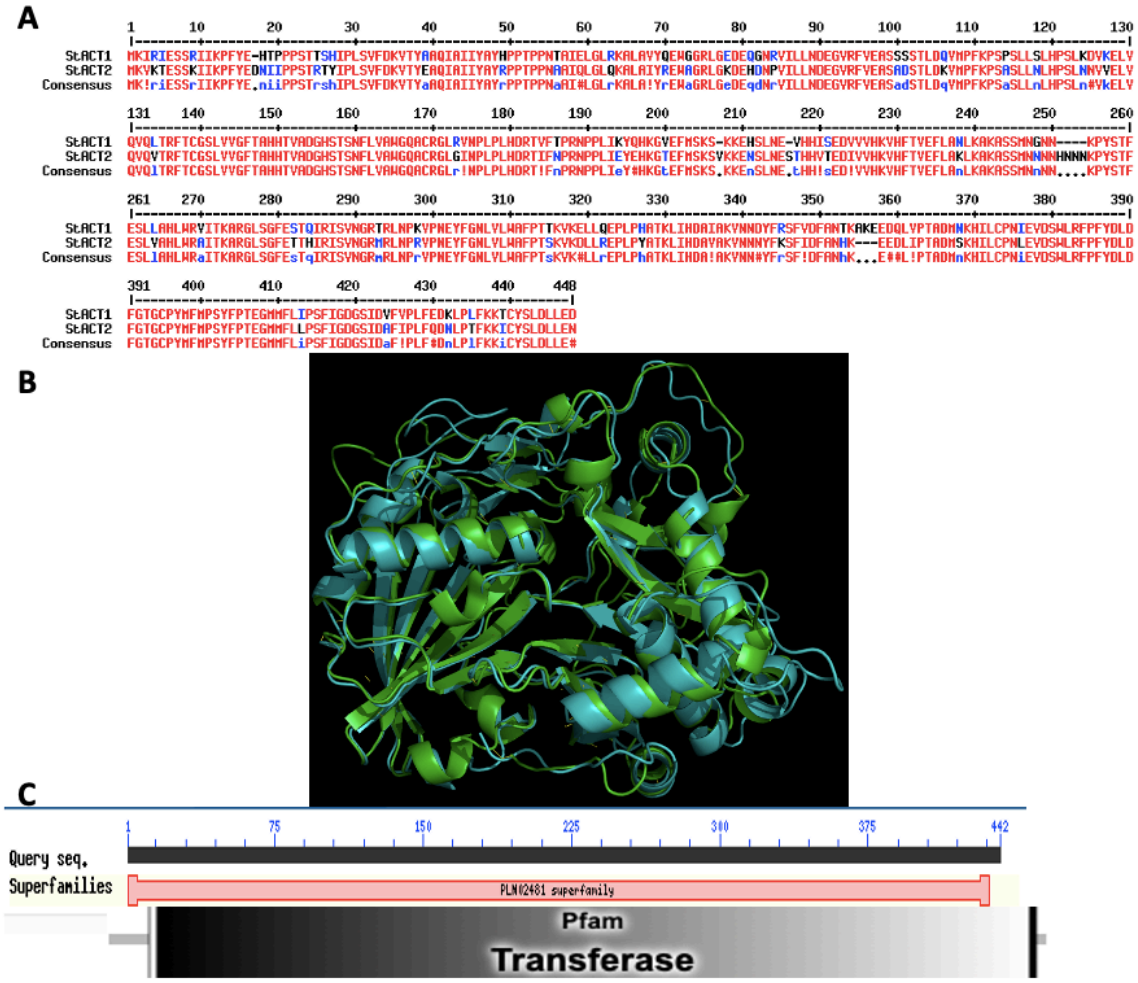


Fig 4.3 Comparison of *StACT1* and *StACT2* and domain identification.

Comparison of Protein sequence and folding between *StACT1* and *StACT2.* A) The amino acid sequence comparison between the two genes and red colored letters shows the identical amino acid at the specific position and the blue shows the unidentical amino acid. The black letters show the gap in the sequence. Some sequence difference can be seen between genes, and also 4 amino acids missing in the *StACT1* at position 251, and 3 amino acids missing at the position 355 in *StACT2*. B) Shows the protein 3D structure superimposed to compare the structural difference in the protein. It shows that no major difference in the secondary structure is seen between *StACT1* and *StACT2* due to the difference in the sequence. C) To identify if the two genes show the same domains and motifs. It shows both the genes show the same transferase domain and no domain difference can be seen.

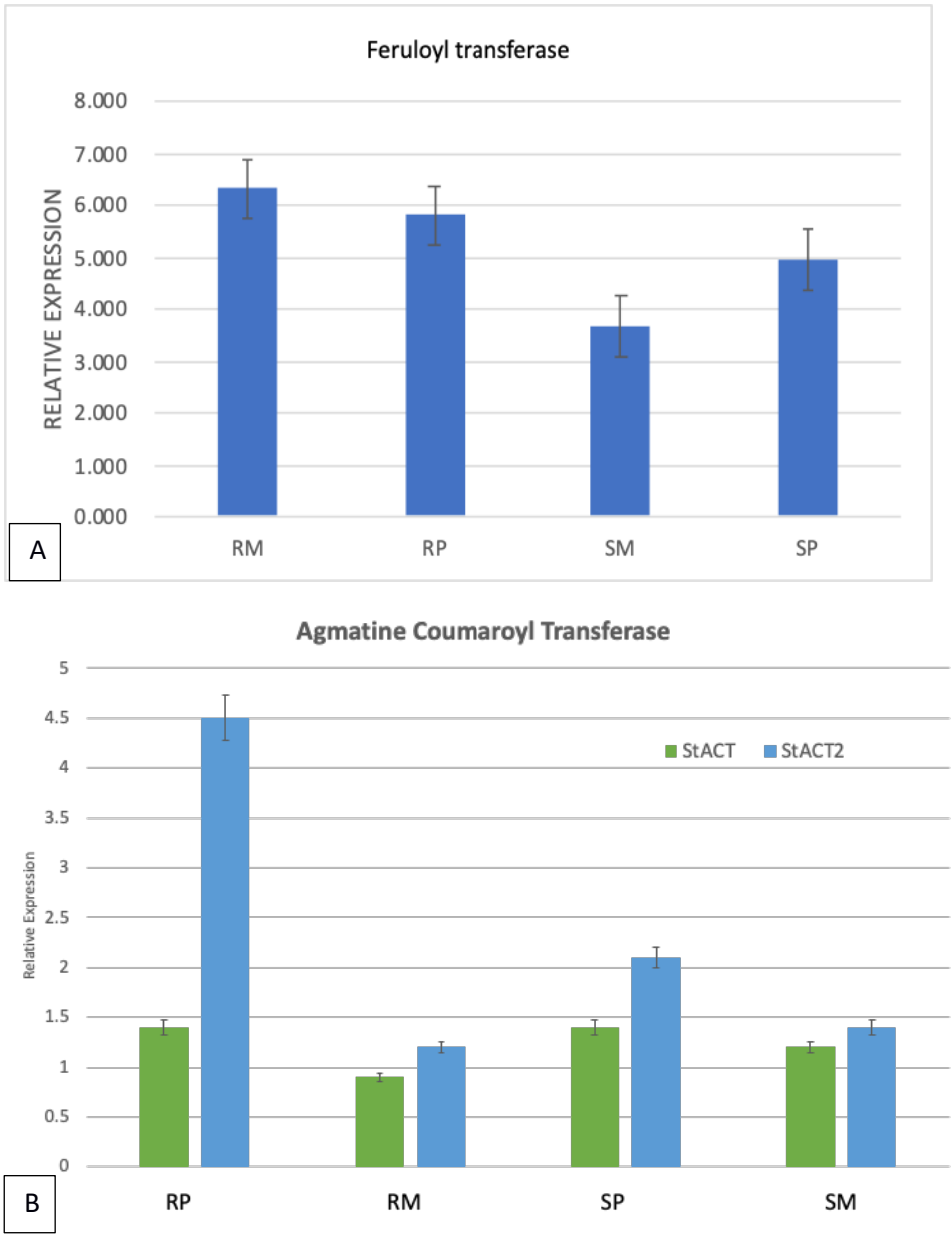


Fig 4.4 Quantitative expression of *StFT*, *StACT1* and *StACT2* in 4 treatments.

A) The relative gene expression of feruloyl transferase (*StFT*) gene in 4 treatments, RP, RM, SP and SM. No significant difference is seen between the treatments in gene expression. B) Shows the relative gene expression of the two *ACT* genes *StACT1* and *StACT2* in RP, RM, SP and SM treatments. The *StACT2* gene shows significant difference in gene expression in RP, that is resistant genotype post pathogen inoculation, whereas we do not see significant difference in other treatments.

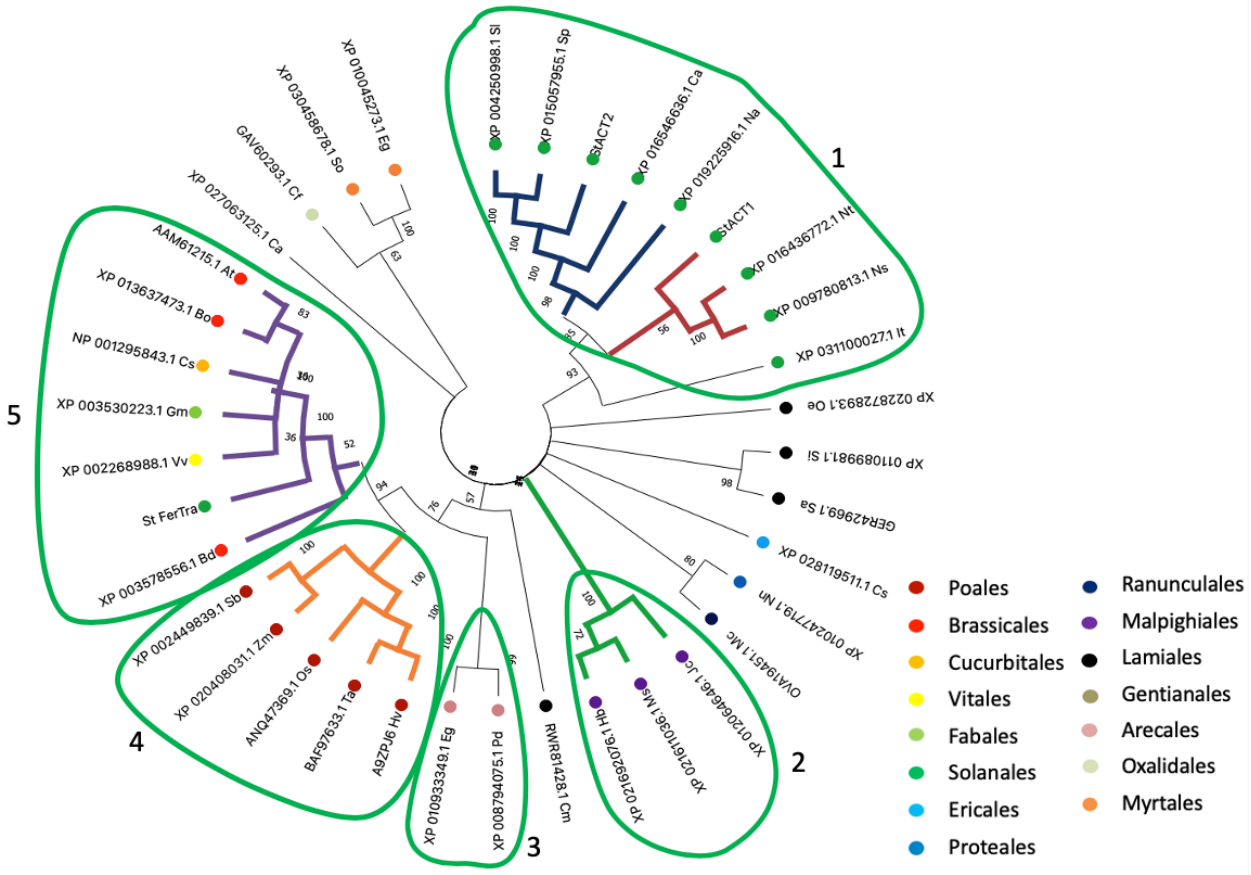


Fig 4.5 Phylogenetic tree depicting the evolutionary linkage to other homologues genes.

The phylogenetic tree depicting the evolutionary relationship between potato *ACT* genes and its closest homologues genes in other plant species. The species are grouped according to the order they belong to and color coded. The 4 important clades are market from 1 to 4. Clade 1 marks both the potato *ACT* genes which shows that there is no significant diversion in between these genes, however, they are grouped under 2 different subclades which denotes certain divergence.

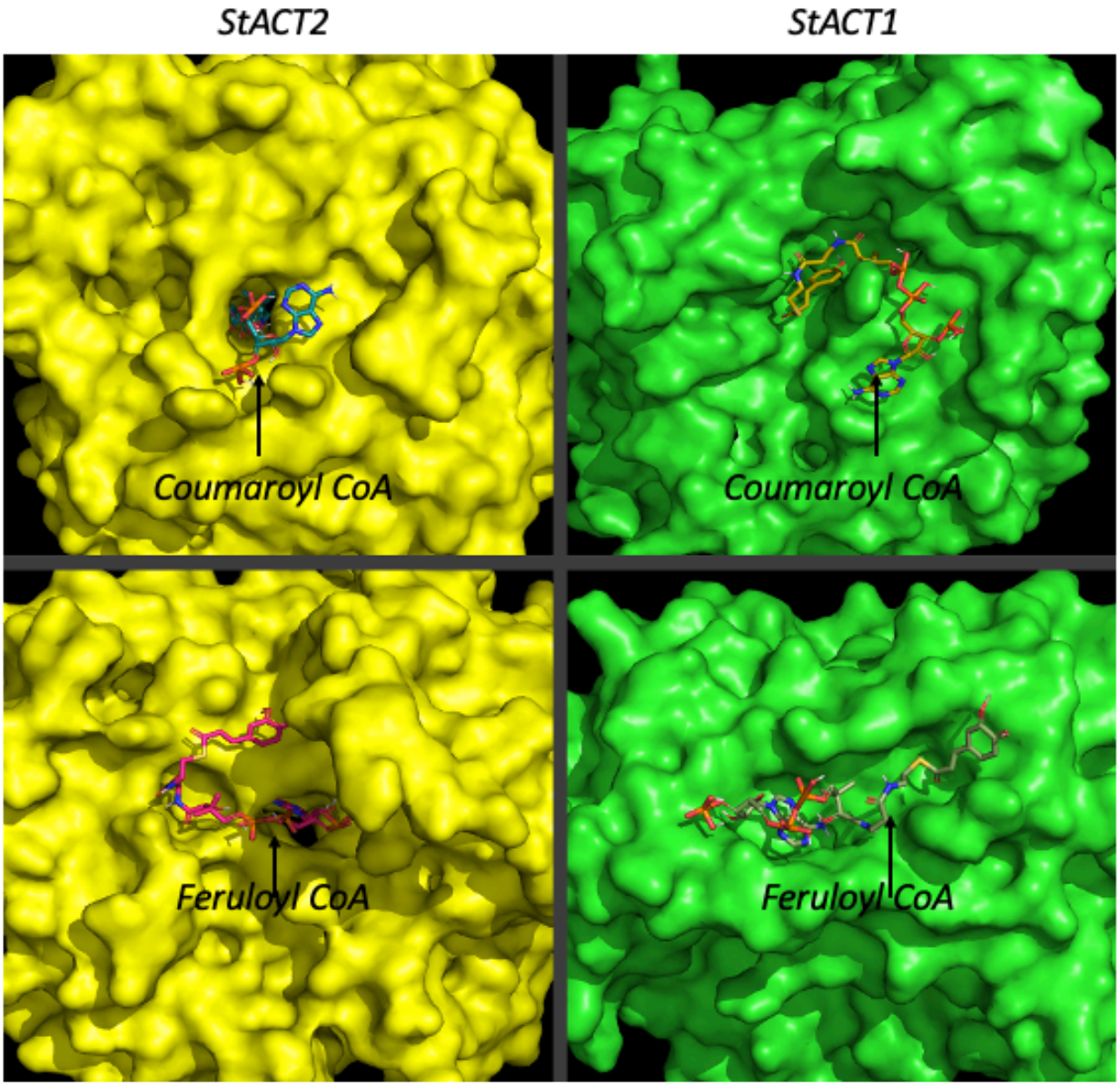


Fig 4.6 Substrate binding to *StACT1* and *StACT2*.

The substrate protein docking studies is depicted comparing the difference in binding between the two ACT genes. The comparison of feruloyl-CoA and coumaroyl-CoA docking to the two proteins *StACT1* and *StACT2* clearly shows that in case of *StACT1* both the substrates does not bind at the active site and the binding is shallow away from the agmatine binding rendering no perfect enzyme activity. Whereas in case of *StACT2* both the substrates bind at the active site which is closer to the agmatine binding facilitating the enzyme activity.

Table 4.2 Binding affinity comparison of substrates between *StACT1* and *StACT2*.

	StACT1		*StACT2*	
Ligand	**Binding affinity coefficient (kcal/mol)**	**RMSD**	**Binding affinity coefficient (kcal/mol)**	**RMSD**
Agmatine	-4.4	0.0	-5.1	0.0
p-coumaroyl-CoA	-8.0	0.0	-9.3	0.0
Feruloyl-CoA	-6.9	0.0	-9.0	0.0

The comparison of the binding affinity coefficient in kcal/mol between the substrate binding in StACT1 and StACT2. It can be seen that there is slight difference in the binding affinity of agmatine whereas there is significant difference in the binding affinity of p-coumaroyl-CoA and feruloyl-CoA to StACT1 and StACT2. The binding is seen stronger in the StACT2 compared to StACT1.

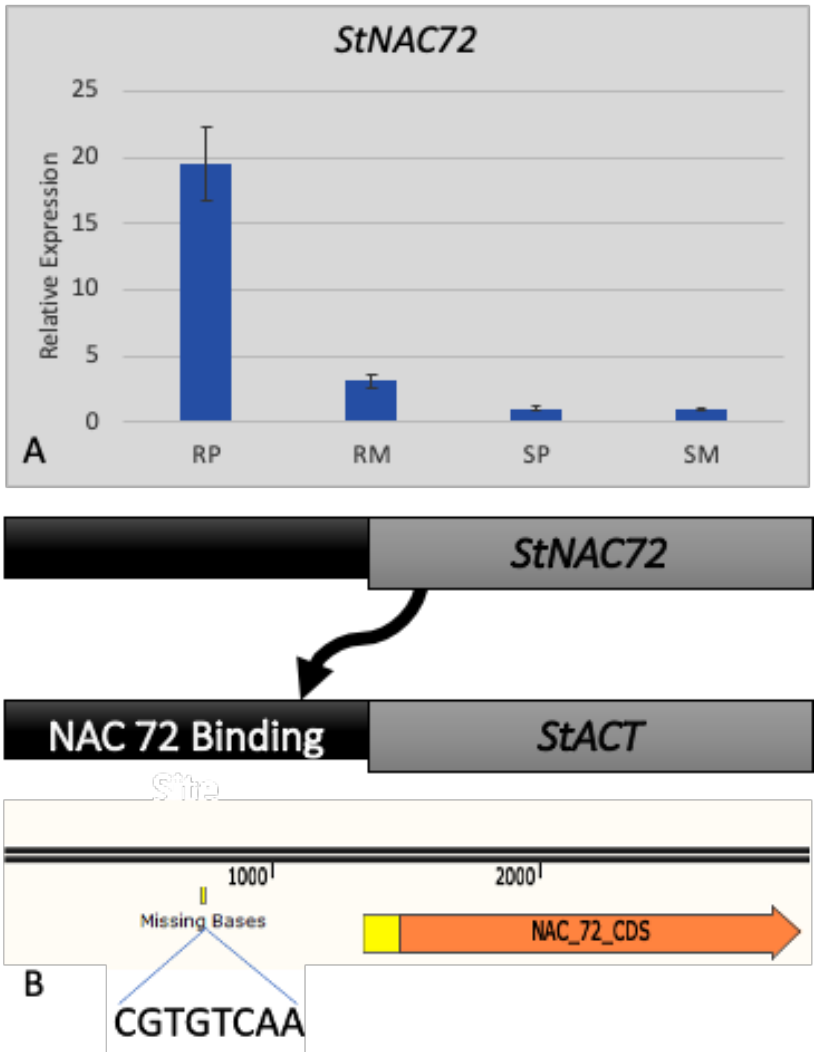


Fig 4.7 *StNAC72* gene expression confirmation and *StNAC-ACT* predicted binding model.

A) Shows the relative gene expression of *StNAC72* in treatments RP, RM, SP and SM. It can be seen that in RP the expression significantly higher (FC=6.3) than RM and we can see no difference in SP and SM. B) The transcription factor binding model and the mutation identified in the *NAC72* promoter in AG704

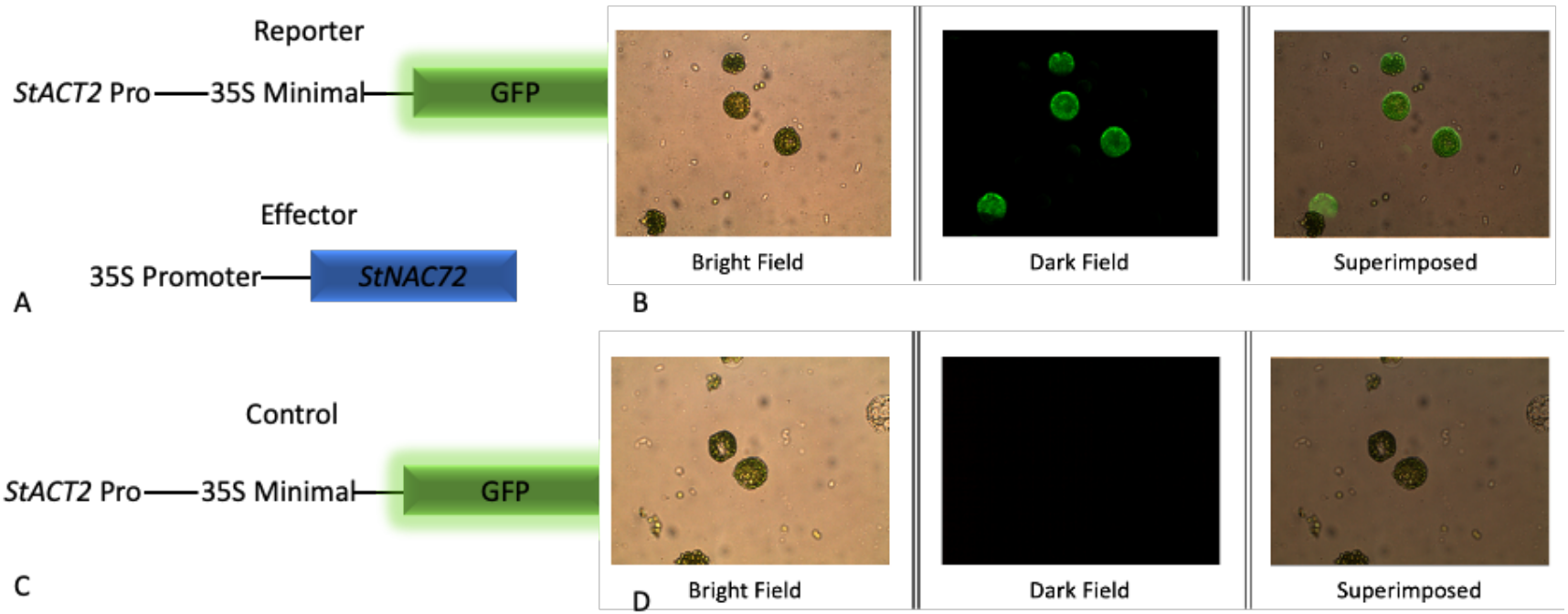


Fig 4.8 GFP assay to find the *in vivo* binding of *StNAC72* to *StACT2* promoter.

A) Shows the reporter and effector model used to identify the *StNAC72* to *StACT2* binding. B) Shows the fluorescence microscopy image of potato protoplast in bright field, dark field and the superimposed image. The fluorescence is seen in the dark field. C) Shows the control model used for the binding assay with no effector. D) Shows the fluorescence microscopy image of potato protoplast in bright field, dark field and the superimposed image. The fluorescence is not seen in the dark field for the control as seen for the test samples in B.

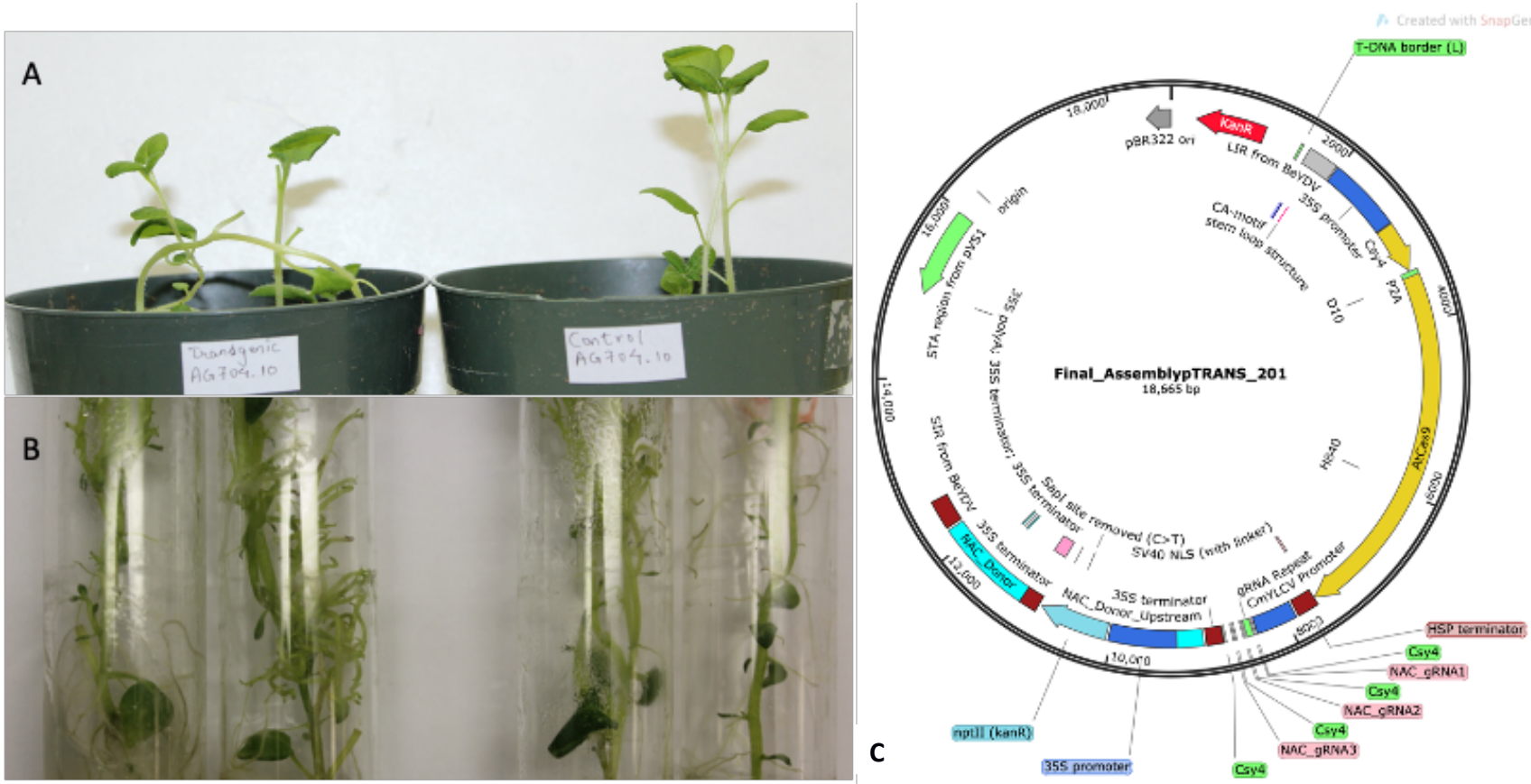


Fig 4.9 Genome editing to generate transgenic AG704 plants

The *StACT2* gene edited (TransAG) (A) and control plants (B) growing in MS medium and transferred to soil. There were no visible phenotypic differences between the edited and control AG704 plants. C) The final destination vector with CRISPR gRNA, Cas9 and NAC donor cloned in pTRANS_201 which has the T-DNA and the BeYDV viral components.

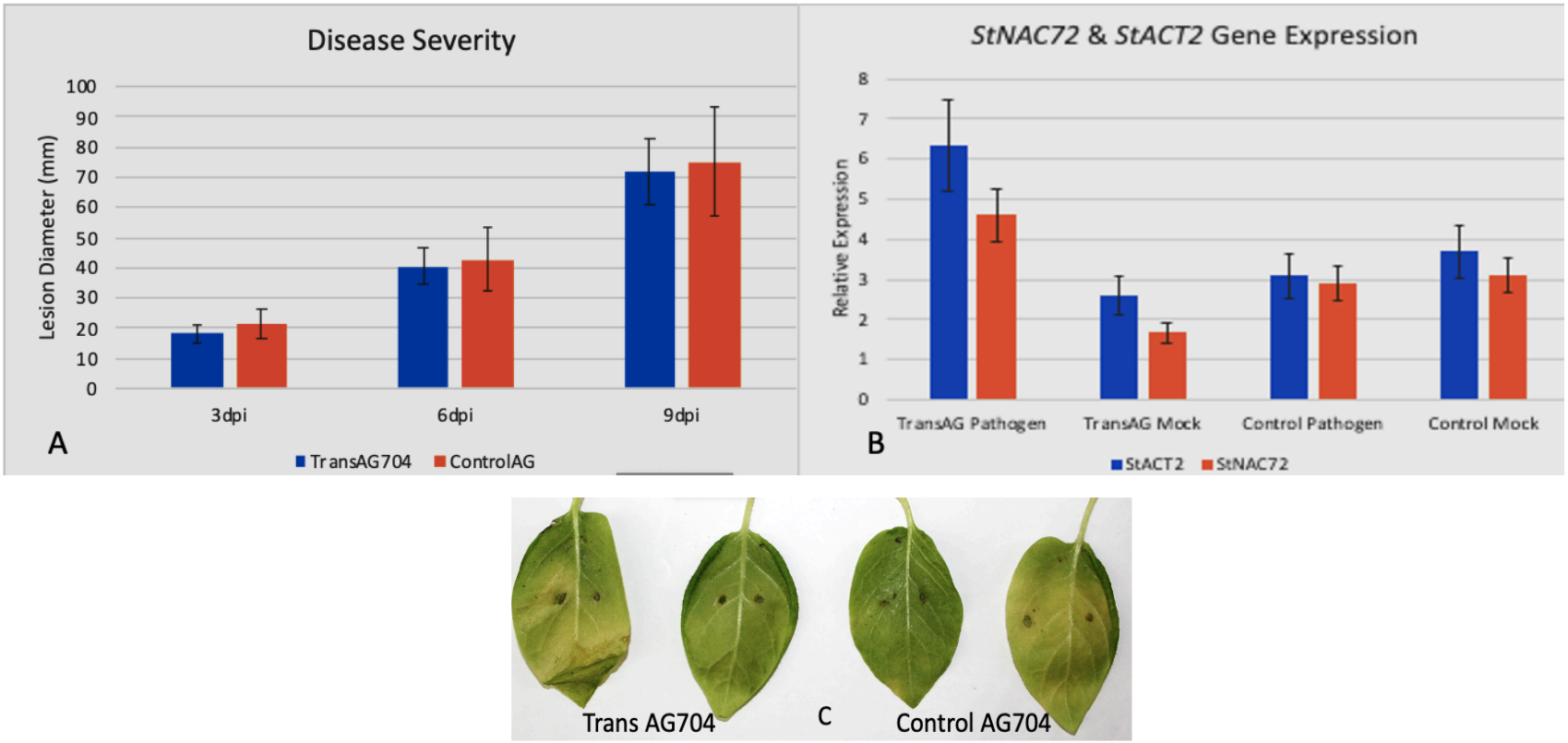


Fig 4.10 Disease severity and gene expression in Trans AG704 and Control AG704.

The disease severity (A and C), lesions, in edited (Trans AG704) and Control (AG704) detached leaves, inoculated with P. infestans. The lesion diameter quantified on 3, 6, and 9 dpi were not significant between the edited and non-edited plants. B The relative gene expression of *StNAC72* and *StACT2* genes in control AG704 and trans AG704 with mock and pathogen inoculation. Both the genes have higher expression in Trans AG704 than in control, post pathogen inoculation.

Reference list

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. Nat Protoc. doi: 10.1038/nprot.2006.384

Bent AF, Mackey D (2007) Elicitors, Effectors, and *R* Genes: The New Paradigm and a Lifetime Supply of Questions. Annu Rev Phytopathol **45**: 399–436

De Block M (1988) Genotype-independent leaf disc transformation of potato (Solanum tuberosum) using Agrobacterium tumefaciens. Theor Appl Genet. doi: 10.1007/BF00303524

Bollina V, Kumarswamy GK, Kushalappa AC, Choo TM, Dion Y, Rioux S, Faubert D, Hamzehzarghan H (2010) Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley against Fusarium head blight. Mol Plant Pathol **11**: 76–82

Burhenne K, Kristensen BK, Rasmussen SK (2003) A new class of N-hydroxycinnamoyltransferases: Purification, cloning, and expression of a barley agmatine coumaroyltransferase (EC 2.3.1.64). J Biol Chem. doi: 10.1074/jbc.M213041200

Carere J, Powell J, Fitzgerald T, Kazan K, Gardiner DM (2018) BdACT2a encodes an agmatine coumaroyl transferase required for pathogen defence in Brachypodium distachyon. Physiol Mol Plant Pathol **104**: 69–76

Čermák T, Curtin SJ, Gil-Humanes J, Čegan R, Kono TJY, Konečná E, Belanto JJ, Starker CG, Mathre JW, Greenstein RL, et al (2017) A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. Plant Cell **29**: 1196–1217

DeLano WL (2002) The PyMOL Molecular Graphics System. Schrödinger LLC wwwpymolorg **Version 1.**: http://www.pymol.org

Du M, Zhao J, Tzeng DTW, Liu Y, Deng L, Yang T, Zhai Q, Wu F, Huang Z, Zhou M, et al (2017) MYC2 orchestrates a hierarchical transcriptional cascade that regulates jasmonate-mediated plant immunity in tomato. Plant Cell **29**: 1883–1906

Engler C, Marillonnet S (2014) Golden Gate cloning. Methods Mol Biol. doi: 10.1007/978-1-62703-764-8_9

Jin S, Yoshida M (2000) Antifungal compound, feruloylagmatine, induced in winter wheat exposed to a low temperature. Biosci Biotechnol Biochem. doi: 10.1271/bbb.64.1614

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-) **337**: 816 LP – 821

Kage U, Karre S, Kushalappa AC, McCartney C (2017a) Identification and characterization of a fusarium head blight resistance gene TaACT in wheat QTL-2DL. Plant Biotechnol J. doi: 10.1111/pbi.12641

Kage U, Yogendra KN, Kushalappa AC (2017b) TaWRKY70 transcription factor in wheat QTL-2DL regulates downstream metabolite biosynthetic genes to resist Fusarium graminearum infection spread within spike. Sci Rep. doi: 10.1038/srep42596

Kushalappa AC, Gunnaiah R (2013) Metabolo-proteomics to discover plant biotic stress resistance genes. Trends Plant Sci **18**: 522–531

Kushalappa AC, Yogendra KN, Karre S (2016) Plant Innate Immune Response: Qualitative and Quantitative Resistance. CRC Crit Rev Plant Sci **35**: 38–55

Li J, Zhang K, Meng Y, Hu J, Ding M, Bian J, Yan M, Han J, Zhou M (2018) Jasmonic acid/ethylene signaling coordinates hydroxycinnamic acid amides biosynthesis through ORA59 transcription factor. Plant J **95**: 444–457

Macoy DM, Kim WY, Lee SY, Kim MG (2015) Biotic stress related functions of hydroxycinnamic acid amide in plants. J Plant Biol. doi: 10.1007/s12374-015-0104-y

Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, Olson AJ (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem **30**: 2785–91

Muroi A, Ishihara A, Tanaka C, Ishizuka A, Takabayashi J, Miyoshi H, Nishioka T (2009) Accumulation of hydroxycinnamic acid amides induced by pathogen infection and identification of agmatine coumaroyltransferase in Arabidopsis thaliana. Planta. doi: 10.1007/s00425-009-0960-0

Onkokesung N, Gaquerel E, Kotkar H, Kaur H, Baldwin IT, Galis I (2012) MYB8 controls inducible phenolamide levels by activating three novel hydroxycinnamoyl-coenzyme A: Polyamine transferases in Nicotiana attenuata. Plant Physiol. doi: 10.1104/pp.111.187229

Peng H, Meyer RS, Yang T, Whitaker BD, Trouth F, Shangguan L, Huang J, Litt A, Little DP, Ke H, et al (2019) A novel hydroxycinnamoyl transferase for synthesis of hydroxycinnamoyl spermine conjugates in plants. BMC Plant Biol. doi: 10.1186/s12870-019-1846-3

Petersen M (2016) Hydroxycinnamoyltransferases in plant metabolism. Phytochem Rev.

doi: 10.1007/s11101-015-9417-1

Pushpa D, Yogendra KN, Gunnaiah R, Kushalappa AC, Murphy A (2013) Identification of Late Blight Resistance-Related Metabolites and Genes in Potato through Nontargeted Metabolomics. Plant Mol Biol Report **32**: 584–595

Soler M, Verdaguer R, Fernández-Piñán S, Company-Arumí D, Boher P, Góngora-Castillo E, Valls M, Anticó E, Molinas M, Serra O, et al (2020) Silencing against the conserved NAC domain of the potato StNAC103 reveals new NAC candidates to repress the suberin associated waxes in phellem. Plant Sci **291**: 110360

Trott O, Olson AJ (2009) AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem **31**: NA-NA

Tsuda K, Somssich IE (2015) Transcriptional networks in plant immunity. New Phytol **206**: 932–947

Voynikov Y, Zheleva-Dimitrova D, Gevrenova R, Lozanov V, Zaharieva MM, Tsvetkova I, Najdenski H, Yagi S, Almoulah NF, Momekov G (2016) Hydroxycinnamic acid amide profile of Solanum schimperianum Hochst by UPLC-HRMS. Int J Mass Spectrom. doi: 10.1016/j.ijms.2016.08.008

Webb B, Sali A (2016) Comparative Protein Structure Modeling Using MODELLER. Curr. Protoc. Bioinforma. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 5.6.1-5.6.37

Yogendra KN, Kumar A, Sarkar K, Li Y, Pushpa D, Mosa KA, Duggavathi R, Kushalappa AC (2015a) Transcription factor StWRKY1 regulates phenylpropanoid metabolites conferring late blight resistance in potato. J Exp Bot **66**: 7377–89

Yogendra KN, Kushalappa AC, Sarmiento F, Rodriguez E, Mosquera T (2015b) Metabolomics deciphers quantitative resistance mechanisms in diploid potato clones against late blight. Funct Plant Biol **42**: 284–298

Yogendra KN, Pushpa D, Mosa KA, Kushalappa AC, Murphy A, Mosquera T (2014) Quantitative resistance in potato leaves to late blight associated with induced hydroxycinnamic acid amides. Funct Integr Genomics **14**: 285–298

Yoo S-D, Cho Y-H, Sheen J (2007) Arabidopsis mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc **2**: 1565–1572

Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V (2017) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. J Mol Biol. doi: 10.1016/J.JMB.2017.12.007

Chapter V

General Discussion, Conclusion, Summary and Future Research

P. infestans is a notorious pathogen which has the potential to disrupt the established potato supply chain by causing the epidemic like in the past. The pathogen physiology and genetic flexibility is an advantage for the invading pathogen to further spread and intrude large areas and cause severe destruction. Currently using fungicides to control late blight is the widely used method. However, understanding the risks of using chemicals at higher rate would add to the environmental problems which the world is already dealing with. Hence, using genetic resistance would be the ideal way to deal with plant pathogens in the future including *P. infestans*. Thus, we designed chapter III and IV to identify the genetic framework which are involved in controlling the resistance mechanism in resistant genotype which are of less economic value.

In chapter III, several induced genes were identified, which biosynthesize resistance related (RR) induced metabolites upon pathogen attack, in higher fold change in a resistant genotype than in a susceptible genotype. The pathogen induced gene expressions and metabolites biosynthesized were obtained by RNAseq and metabolomics, respectively. A total of 281 constitutive and 160 induced metabolites, belonging to different chemical groups were identified. RNAseq de novo assembly revealed 611 induced genes which were further categorized based on their biological functions. Promoter sequences of these genes were identified using in house script, which was used for Transcription Factor (TF) enrichment study to identify regulatory TFs. A total of 134 TFs were highly enriched in the promoters of induced genes and a correlation study among the induced genes and the enriched TFs further mapped several primary and secondary TFs. Among these *bHLH66, MYB61, NAC56, WRKY51*, *MYB* like, *ERF RAP2-3* and *MADS-box AGL15* had regulating sites in more than 208 downstream genes, of which many were secondary TFs. Hence, a two-tier transcriptional regulation of defense response genes was mapped. This study offers a deeper insight into the complex regulatory role of genes and their biosynthetic metabolites in disease resistance. Following validation, these genes can be used to develop disease resistant cultivars.

One class of secondary metabolites have shown to work effectively to contain several pathogens in plants including Arabidopsis, barley, wheat, and others. These are the hydroxycinnamic acid amides (HCAAs). The HCAAs from the previous metabolomic

experiment were identified and analysed to find the disease resistance response related HCAAs. We identified 4 specific types of amines namely, agmatines, cadavarines, Putrescines and Tyramines from the metabolomics data, however only feruloylagmatine was highly accumulated in the resistant genotype post pathogen inoculation. To further characterize the genes responsible for the accumulation of feruloylagmatine, transferase genes from potato were identified. We analysed 2 major genes which might be responsible and were named according to the nomenclature as *StACT1* and *StACT2*. Along with these genes another gene that is Feruloyl transferase gene was also analysed to find if it can be catalysing the feruloylagmatine formation. Analysing the gene sequencing, qPCR and *in silico* protein docking data showed us that *StACT*2 gene should be responsible for the feruloylagmatine accumulation in potato.

We could identify that *StACT2* is the gene responsible for catalysing feruloylagmatine, however, it was unclear as what is the difference in the susceptible and resistant genotype that is responsible for differential accumulation. We could not find any major sequence difference in the CDS region and hence we sequenced the promoter region which had no sequence difference either. However, we found a *NAC* transcription factor binding site in the *StACT2* promoter region and we could identify that *NAC72* TF was differentially accumulated in susceptible and resistance genotype. We further conducted transcription factor binding assay in protoplast which showed that *NAC72* promotes the transcription of *StACT2*. When the *NAC72* CDS and promoter was sequenced we found a mutation in the promoter of *NAC72*. We further conducted a genome editing experiment with the susceptible genotype to identify if correcting this mutation can further enhance the *StACT2* gene. Transgenic lines were generated using CRISPR Cas9 genome editing and positive transformants were selected which had the missing sequence inserted. These transgenic lines were further inoculated with *P. infestans* spores along with control plants. We found increase in the *StACT2* and *NAC72* gene expression in the transgenic plants, however disease severity showed no significant difference. Other genes involved in the regulation of StACT2 must be explored to enhance feruloylagmatine.

The importance of functional genomics is not only the identification of gene functions but also the downstream application of the functional knowledge to improve the crop varieties. Marker assisted selection (MAS) is one of the most important technique used in breeding programs for plant improvement. However, these markers, even when they are fine mapped, contain several genes, including undesirable genes, thus breeding based on MAS is very challenging. Even when a gene specific marker is involved the segregation would be a limiting factor, thus again transferring several genes, including undesirable genes (Kushalappa et al., 2016b). The

development of commercial cultivars through breeding has been mainly focused on increasing yield, and thus their genetic diversity for biotic stress resistance is rather low. These cultivars are expected to have several functional genes in the cascade or cascades but a few of them may be non-functional, thus unable to biosynthesize a given set of metabolites, rendering the cultivar susceptible. Replacing these non-functional genes with functional segments, it would be possible to improve resistance of the cultivar to late blight.

With the advent of genome editing techniques like CRISPR-Cas9, molecular breeding is becoming a feasible option for plant improvement. The resistance in plants against pathogen attack depends on the additive effects of several genes that biosynthesize RR proteins and RR metabolites (Kushalappa et al., 2016a). The resistance as a phenotype is the end-product of all the cascades of resistance genes and thus in the susceptible genotype if one cascade is made functional by replacing/introducing the mutated genetic factors, the resistance can be relatively increased.

Suggested future studies

- Identifying the transcription factor mapping of several potato genotypes to identify the consistent common factors which appear regulating the resistance.
- Metabolomics and transcriptomics and complementing it with TF mapping of other plant pathogen systems to identify the factors that can confer multiple disease resistance.
- The genes, metabolites and transcription factors identified in this study can be useful to develop commercial genotypes which are more valuable and ecofriendly after further validation.
- Similar maps like in this study, can be further extrapolated to other related plant species in the Solanaceae family like tomato.
- Several high impact transcription factors were identified which can be experimentally validated in potato and other crops to identify their impact.
- Metabolites, genes and TFs identified in this study can be further explored experimentally in these genotypes to understand the resistance mechanism and other missing links in the susceptible genotype.
- Metabolomics of the transgenic lines generated could further help in understanding the feruloylagmatine pathway and also other pathways that were affected.

- Validating primary and secondary TFs identified in this study further through *in vivo* studies could help in understanding their specific their role in mounting resistance response.
- TFs identified in this study could be useful for the identification of *in vivo* target analysis like chromatin immunoprecipitation (ChIP) which can help in understanding their role in greater details.
- The TFs can also be useful for projects like ENCODE (Encyclopedia of DNA Elements), which have not yet been formulated.

Reference List

Agrios G (2005) Plant Pathology 5th Edition. San Diego Acad Press 922

Ahuja I, Kissen R, Bones AM (2012) Phytoalexins in defense against pathogens. Trends Plant Sci **17**: 73–90

Anandalakshmi R, Pruss GJ, Ge X, Marathe R, Mallory AC, Smith TH, Vance VB (1998) A viral suppressor of gene silencing in plants. Proc Natl Acad Sci U S A **95**: 13079–84

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

Baur M, Potrykus I, Paszkowski J (1990) Intermolecular homologous recombination in plants. Mol Cell Biol **10**: 492–500

Benton HP, Wong DM, Trauger SA, Siuzdak G (2008) XCMS ^{2} : Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. Anal Chem **80**: 6382–6389

Bhaskar PB, Venkateshwaran M, Wu L, Ané JM, Jiang J (2009) Agrobacterium-mediated transient gene expression and silencing: A rapid tool for functional gene assay in potato. PLoS One. doi: 10.1371/journal.pone.0005812

Birch PRJ, Bryan G, Fenton B, Gilroy EM, Hein I, Jones JT, Prashar A, Taylor MA, Torrance L, Toth IK (2012) Crops that feed the world 8: Potato: Are the trends of increased global production sustainable? Food Secur **4**: 477–508

Bollina V, Kumarswamy GK, Kushalappa AC, Choo TM, Dion Y, Rioux S, Faubert D, Hamzehzarghan H (2010) Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley against Fusarium head blight. Mol Plant Pathol **11**: 76–82

Bollina V, Kushalappa AC, Choo TM, Dion Y, Rioux S (2011) Identification of metabolites related to mechanisms of resistance in barley against Fusarium graminearum, based on

mass spectrometry. Plant Mol Biol **77**: 355–370

Bradshaw JE, Bryan GJ, Lees AK, McLean K, Solomon-Blackburn RM (2006) Mapping the R10 and R11 genes for resistance to late blight (Phytophthora infestans) present in the potato (Solanum tuberosum) R-gene differentials of Black. Theor Appl Genet **112**: 744–751

Čermák T, Curtin SJ, Gil-Humanes J, Čegan R, Kono TJY, Konečná E, Belanto JJ, Starker CG, Mathre JW, Greenstein RL, et al (2017) A Multipurpose Toolkit to Enable Advanced Genome Engineering in Plants. Plant Cell **29**: 1196–1217

Chakrabarti SK, Singh BP, Thakur G, Tiwari JK, Kaushik SK, Sharma S, Bhardwaj V (2014) QTL Analysis of Late Blight Resistance in a Diploid Potato Family of Solanum spegazzinii × S. chacoense. Potato Res **57**: 1–11

Champouret N, Bouwmeester K, Rietman H, van der Lee T, Maliepaard C, Heupink A, van de Vondervoort PJI, Jacobsen E, Visser RGF, van der Vossen EAG, et al (2009) *Phytophthora infestans* Isolates Lacking Class I *ipiO* Variants Are Virulent on *Rpi-blb1* Potato. Mol Plant-Microbe Interact **22**: 1535–1545

Chen C-H, Sheu Z-M, Wang T-C (2008) Host Specificity and Tomato-Related Race Composition of Phytophthora infestans Isolates in Taiwan During 2004 and 2005. Plant Dis **92**: 751–755

Chisholm ST, Coaker G, Day B, Staskawicz BJ (2006) Host-Microbe Interactions: Shaping the Evolution of the Plant Immune Response. Cell **124**: 803–814

Chowdappa P, Kumar NBJ, Madhura S, Kumar MSP, Myers KL, Fry WE, Squires JN, Cooke DEL (2013) Emergence of 13_A2 Blue Lineage of *Phytophthora infestans* was Responsible for Severe Outbreaks of Late Blight on Tomato in South-West India. J Phytopathol **161**: 49–58

Colin A. Smith, Elizabeth J. Want, Grace O'Maille, Ruben Abagyan A, Siuzdak* G (2006) XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Anal Chem **78**: 779–787

Collins A, Milbourne D, Ramsay L, Meyer R, Chatot-Balandras C, Oberhagemann P, De

Jong W, Gebhardt C, Bonnel E, Waugh R (1999) QTL for field resistance to late blight in potato are strongly correlated with maturity and vigour. Mol Breed **5**: 387–398

Danan S, Veyrieras J-B, Lefebvre V (2011) Construction of a potato consensus map and QTL meta-analysis offer new insights into the genetic architecture of late blight resistance and plant maturity traits. BMC Plant Biol **11**: 16

Dangl JL, Jones JDG (2001) Plant pathogens and integrated defence responses to infection. Nature **411**: 826–833

Engler C, Marillonnet S (2014) Golden Gate cloning. Methods Mol Biol. doi: 10.1007/978-1-62703-764-8_9

Eudes F, Chugh A (2008) Cell-penetrating peptides. Plant Signal Behav **3**: 549–550

Feng Z, Zhang B, Ding W, Liu X, Yang D-L, Wei P, Cao F, Zhu S, Zhang F, Mao Y, et al (2013) Efficient genome editing in plants using a CRISPR/Cas system. Cell Res **23**: 1229–1232

Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Mol Biol **48**: 155–171

Fiehn O, Kopka J, Dörmann P, Altmann T, Trethewey RN, Willmitzer L (2000) Metabolite profiling for plant functional genomics. Nat Biotechnol **18**: 1157–1161

Filippov A, Rogozhin A, Kuznetsova M, Statsyuk N, Ronis A, Platt HW (Bud) (2015) Efficiency of a computerised decision support system as a tool to reduce fungicide input for the control of potato late blight. Zemdirbyste-Agriculture **102**: 449–456

Fisher MC, Henk D a, Briggs CJ, Brownstein JS, Madoff LC, McCraw SL, Gurr SJ (2012) Emerging fungal threats to animal, plant and ecosystem health. Nature **484**: 186–94

Flor HH (1971) Current status of the Gene-For-Gene concept. Annu Rev Phytopathol **9**: 275–296

Freund DM, Hegeman AD (2017) Recent advances in stable isotope-enabled mass spectrometry-based plant metabolomics. Curr Opin Biotechnol **43**: 41–48

Fry W (2008) Phytophthora infestans: The plant (and R gene) destroyer. Mol Plant Pathol **9**: 385–402

Fry WE (1977) Integrated Control of Potato Late Blight -Effects of Polygenic Resistance and Techniques of Timing Fungicide Applications. Phypathology **67**: 415–420

Fry WE (2016) *Phytophthora infestans*: New Tools (and Old Ones) Lead to New Understanding and Precision Management. Annu Rev Phytopathol **54**: 529–547

Fry WE, Birch PRJ, Judelson HS, Grünwald NJ, Danies G, Everts KL, Gevens AJ, Gugino BK, Johnson DA, Johnson SB, et al (2015) Five Reasons to Consider *Phytophthora infestans* a Reemerging Pathogen. Phytopathology **105**: 966–981

Fry WE, Goodwin SB (1997) Reemergence of potato and tomato late bligth in United States. Plant Desease 69–72

Fry WE, McGrath MT, Seaman a., Zitter T a., McLeod a., Danies G, Small IM, Myers K, Everts K, Gevens a. J, et al (2013) The 2009 late blight pandemic in Eastern USA – causes and results. Plant Dis **97**: 296–306

Gil-Humanes J, Wang Y, Liang Z, Shan Q, Ozuna C V., Sánchez-León S, Baltes NJ, Starker C, Barro F, Gao C, et al (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. Plant J **89**: 1251–1262

Gisi U, Cohen Y (1996) RESISTANCE TO PHENYLAMIDE FUNGICIDES:A Case Study with *Phytophthora infestans* Involving Mating Type and Race Structure. Annu Rev Phytopathol **34**: 549–572

Gisi U, Walder F, Resheat-Eini Z, Edel D, Sierotzki H (2011) Changes of Genotype, Sensitivity and Aggressiveness in Phytophthora infestans Isolates Collected in European Countries in 1997, 2006 and 2007. J Phytopathol **159**: 223–232

Gorbunova V, Levy AA (1997) Non-homologous DNA end joining in plant cells is associated with deletions and filler DNA insertions. Nucleic Acids Res **25**: 4650–4657

Goss EM, Tabima JF, Cooke DEL, Restrepo S, Fry WE, Forbes GA, Fieland VJ, Cardenas M, Grünwald NJ (2014) The Irish potato famine pathogen Phytophthora infestans originated in central Mexico rather than the Andes. Proc Natl Acad Sci U S A

111: 8791–6

Gunnaiah R, Kushalappa AC, Duggavathi R, Fox S, Somers DJ, Young N, Danan S, Veyrieras J, Lefebvre V, Ballini E, et al (2012) Integrated Metabolo-Proteomic Approach to Decipher the Mechanisms by Which Wheat QTL (Fhb1) Contributes to Resistance against Fusarium graminearum. PLoS One **7**: 479–501

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM, Grabherr M, Kodira CD, Raffaele S, Torto-Alalibo T, et al (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature **461**: 393–398

Hall R, Beale M, Fiehn O, Hardy N, Sumner L, Bino R (2002) Plant metabolomics: the missing link in functional genomics strategies. Plant Cell **14**: 1437–40

Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytol **169**: 453–468

Hall RD (2018) Plant Metabolomics in a Nutshell: Potential and Future Challenges. Annu. Plant Rev. online. John Wiley & Sons, Ltd, Chichester, UK, pp 1–24

Hamzehzarghani H, Kushalappa AC, Dion Y, Rioux S, Comeau A, Yaylayan V, Marshall WD, Mather DE (2005) Metabolic profiling and factor analysis to discriminate quantitative resistance in wheat cultivars against fusarium head blight. Physiol Mol Plant Pathol **66**: 119–133

Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009) Applied biotechnology to combat late blight in potato caused by phytophthora infestans. Potato Res **52**: 249–264

Hijmans RJ, Forbes GA, Walker TS (2000) Estimating the global severity of potato late blight with GIS-linked disease forecast models. Plant Pathol **49**: 697–705

Hirsch CD, Hamilton JP, Childs KL, Cepela J, Crisovan E, Vaillancourt B, Hirsch CN, Habermann M, Neal B, Buell CR (2014) Spud DB: A Resource for Mining Sequences, Genotypes, and Phenotypes to Accelerate Potato Breeding. Plant Genome **7**: 0

Hu X, Wang C, Liu Q, Fu Y, Wang K (2017) Targeted mutagenesis in rice using CRISPR-Cpf1 system. J Genet Genomics **44**: 71–73

Huang S, Weigel D, Beachy RN, Li J (2016) A proposed regulatory framework for genome-edited crops. Nat Genet **48**: 109–111

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science (80-) **337**: 816 LP – 821

Jo K-R, Kim C-J, Kim S-J, Kim T-Y, Bergervoet M, Jongsma MA, Visser RG, Jacobsen E, Vossen JH, Muller K, et al (2014) Development of late blight resistant potatoes by cisgene stacking. BMC Biotechnol **14**: 50

Karunasagar I, Karunasagar I (2016) Challenges of Food Security – Need for Interdisciplinary Collaboration. Procedia Food Sci **6**: 31–33

Kaspar S, Peukert M, Svatos A, Matros A, Mock H-P (2011) MALDI-imaging mass spectrometry - An emerging technique in plant biology. Proteomics **11**: 1840–1850

Katajamaa M, Miettinen J, Orešič M (2006) MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics **22**: 634–636

Klein M, Eslami-Mossallam B, Arroyo DG, Depken M (2018) Hybridization Kinetics Explains CRISPR-Cas Off-Targeting Rules. Cell Rep **22**: 1413–1423

Kumaraswamy GK, Bollina V, Kushalappa AC, Choo TM, Dion Y, Rioux S, Mamer O, Faubert D (2011a) Metabolomics technology to phenotype resistance in barley against Gibberella zeae. Eur J Plant Pathol **130**: 29–43

Kumaraswamy KG, Kushalappa AC, Choo TM, Dion Y, Rioux S (2011b) Mass Spectrometry Based Metabolomics to Identify Potential Biomarkers for Resistance in Barley against Fusarium Head Blight (Fusarium graminearum). J Chem Ecol **37**: 846–856

Kushalappa AC, Gunnaiah R (2013) Metabolo-proteomics to discover plant biotic stress resistance genes. Trends Plant Sci **18**: 522–531

Kushalappa AC, Yogendra KN, Karre S (2016a) Plant Innate Immune Response: Qualitative and Quantitative Resistance. CRC Crit Rev Plant Sci **35**: 38–55

Kushalappa AC, Yogendra KN, Sarkar K, Kage U, Karre S (2016b) Gene discovery and genome editing to develop cisgenic crops with improved resistance against pathogen infection. Can J Plant Pathol 07060661.2016.1199597

Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E (2016) CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Res **44**: W272–W276

Łaźniewska J, Macioszek VK, Kononowicz AK (2012) Plant-fungus interface: The role of surface structures in plant resistance and susceptibility to pathogenic fungi. Physiol Mol Plant Pathol **78**: 24–30

Li J-F, Norville JE, Aach J, McCormack M, Zhang D, Bush J, Church GM, Sheen J (2013) Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nat Biotechnol **31**: 688–691

Liang Z, Chen K, Li T, Zhang Y, Wang Y, Zhao Q, Liu J, Zhang H, Liu C, Ran Y, et al (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nat Commun **8**: 14261

Lisec J, Schauer N, Kopka J, Willmitzer L, Fernie AR (2006) Gas chromatography mass spectrometry–based metabolite profiling in plants. Nat Protoc **1**: 387–396

Liu H, Ding Y, Zhou Y, Jin W, Xie K, Chen LL (2017) CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. Mol Plant **10**: 530–532

Lommen A, Kools HJ (2012) MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. Metabolomics **8**: 719–726

van Loon LC, Rep M, Pieterse CMJ (2006) Significance of Inducible Defense-related Proteins in Infected Plants. Annu Rev Phytopathol **44**: 135–162

Van Loon LC, Strien EA Van (1999) The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins. Physiol Mol Plant Pathol **55**: 85–97

MacLeod A, Jones GD, Anderson HM, Mumford RA (2016) Plant health and food security,

linking science, economics, policy and industry. Food Secur **8**: 17–25

Matyssek R, Schnyder H, Elstner E-F, Munch J-C, Pretzsch H, Sandermann H (2002) Growth and Parasite Defence in Plants; the Balance between Resource Sequestration and Retention: In Lieu of a Guest Editorial. Plant Biol **4**: 133–136

Meraj TA, Fu J, Raza MA, Zhu C, Shen Q, Xu D, Wang Q (2020) Transcriptional Factors Regulate Plant Stress Responses Through Mediating Secondary Metabolism. Genes (Basel) **11**: 346

Mosquera T, Alvarez MF, Jiménez-Gómez JM, Muktar MS, Paulo MJ, Steinemann S, Li J, Draffehn A, Hofmann A, Lübeck J, et al (2016) Targeted and Untargeted Approaches Unravel Novel Candidate Genes and Diagnostic SNPs for Quantitative Resistance of the Potato (Solanum tuberosum L.) to Phytophthora infestans Causing the Late Blight Disease. PLoS One **11**: e0156254

Mundt CC, Cowger C, Garrett KA (2002) Relevance of integrated disease management to resistance durability. Euphytica **124**: 245–252

Nawrot R, Barylski J, Nowicki G, Broniarczyk J, Buchwald W, Goździcka-Józefiak A (2014) Plant antimicrobial peptides. Folia Microbiol (Praha) **59**: 181–196

Nekrasov V, Staskawicz B, Weigel D, Jones JDG, Kamoun S (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat Biotechnol **31**: 691–693

Novy RG, Love SL, Corsini DL, Pavek JJ, Whitworth JL, Mosley AR, James SR, Hane DC, Shock CC, Rykbost KA, et al (2006) Defender: A high-yielding, processing potato cultivar with foliar and tuber resistance to late blight. Am J Potato Res **83**: 9–19

Nowicki M, Foolad MR, Nowakowska M, Kozik EU (2012) Potato and Tomato Late Blight Caused by Phytophthora infestans : An Overview of Pathology and Resistance Breeding. Plant Dis **96**: 4–17

Nunn N, Qian N (2011) The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment. Q J Econ **126**: 593–650

Okazaki Y, Saito K (2012) Recent advances of metabolomics in plant biotechnology. Plant

Biotechnol Rep **6**: 1–15

Pennisi E (2013) The CRISPR craze. Science **341**: 833–6

Peters RD, Platt HW (Bud), Hall R (1998) Characterization of changes in populations of *Phytophthora infestans* in Canada using mating type and metalaxyl sensitivity markers. Can J Plant Pathol **20**: 259–273

Piasecka A, Jedrzejczak-Rey N, Bednarek P (2015) Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. New Phytol **206**: 948–964

Pluskal T, Castillo S, Villar-Briones A, Oresic M (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics **11**: 395

Poland JA, Balint-Kurti PJ, Wisser RJ, Pratt RC, Nelson RJ (2009) Shades of gray: the world of quantitative disease resistance. Trends Plant Sci **14**: 21–29

Price T. (2004) Pest and disease incursions: risks, threats and management in Papua New Guinea. Papers presented at the 2nd Papua New Guinea Plant Protection Conference, Kokopo, East New Britain Province. doi: 10.1071/AP05059

Puchta H, Dujon B, Hohn B (1996) Two different but related mechanisms are used in plants for the repair of genomic double-strand breaks by homologous recombination. Proc Natl Acad Sci **93**: 5055–5060

Pushpa D, Yogendra KN, Gunnaiah R, Kushalappa AC, Murphy A (2013) Identification of Late Blight Resistance-Related Metabolites and Genes in Potato through Nontargeted Metabolomics. Plant Mol Biol Report **32**: 584–595

Saville AC, Martin MD, Ristaino JB, Stukenbrock E, McDonald B, Anderson P, Cunningham A, Patel N, Morales F, Epstein P, et al (2016) Historic Late Blight Outbreaks Caused by a Widespread Dominant Lineage of Phytophthora infestans (Mont.) de Bary. PLoS One **11**: e0168381

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes MG, Willmitzer L, et al (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett **579**: 1332–

1337

Schweiger W, Boddu J, Shin S, Poppenberger B, Berthiller F, Lemmens M, Muehlbauer GJ, Adam G (2010) Validation of a Candidate Deoxynivalenol-Inactivating UDP-Glucosyltransferase from Barley by Heterologous Expression in Yeast. Mol Plant-Microbe Interact **23**: 977–986

Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas system. Nat Protoc **9**: 2395–410

Shulaev V (2006) Metabolomics technology and bioinformatics. Brief Bioinform **7**: 128–139

Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. Physiol Plant **132**: 199–208

Simko I (2002) Comparative analysis of quantitative trait loci for foliage resistance to Phytophthora infestans in tuber-bearing Solanum species. Am J Potato Res **79**: 125–132

Slater AT, Cogan NOI, Hayes BJ, Schultz L, Dale MFB, Bryan GJ, Forster JW (2014) Improving breeding efficiency in potato using molecular and quantitative genetics. Theor Appl Genet **127**: 2279–92

Śliwka J, Jakuczun H, Lebecka R, Marczewski W, Gebhardt C, Zimnoch-Guzowska E (2007) Tagging QTLs for late blight resistance and plant maturity from diploid wild relatives in a cultivated potato (Solanum tuberosum) background. Theor Appl Genet **115**: 101–112

Small IM, Joseph L, Fry WE (2015) Evaluation of the BlightPro Decision Support System for Management of Potato Late Blight Using Computer Simulation and Field Validation. Phytopathology **105**: 1545–54

Soga T, Baran R, Suematsu M, Ueno Y, Ikeda S, Sakurakawa T, Kakazu Y, Ishikawa T, Robert M, Nishioka T, et al (2006) Differential metabolomics reveals ophthalmic acid as an oxidative stress biomarker indicating hepatic glutathione consumption. J Biol Chem **281**: 16768–76

Song J, Bradeen JM, Naess SK, Raasch JA, Wielgus SM, Haberlach GT, Liu J, Kuang H, Austin-Phillips S, Buell CR, et al (2003) Gene RB cloned from Solanum

bulbocastanum confers broad spectrum resistance to potato late blight. Proc Natl Acad Sci **100**: 9128–9133

St.Clair DA (2010) Quantitative Disease Resistance and Quantitative Resistance Loci in Breeding. Annu Rev Phytopathol **48**: 247–268

Sujkowski LS, Goodwin SB, Dyer AT, Fry WE (1994) Increased genotypic diversity via migration and possible occurrence of sexual reproduction of Phytophthora infestans in Poland. Phytopathology **84**: 201–207

Thurston HD (1971) Relationship of general resistance - late blight of potato. Phytopathology **61**: 620–626

Tiwari JK, Devi S, Sharma S, Chandel P, Rawat S, Singh BP (2015) Allele Mining in Solanum Germplasm: Cloning and Characterization of RB-Homologous Gene Fragments from Late Blight Resistant Wild Potato Species. Plant Mol Biol Report **33**: 1584–1598

Tohge T, Fernie AR (2010) Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. Nat Protoc **5**: 1210–1227

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc **7**: 562–78

Urano K, Maruyama K, Ogata Y, Morishita Y, Takeda M, Sakurai N, Suzuki H, Saito K, Shibata D, Kobayashi M, et al (2009) Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. Plant J **57**: 1065–1078

De Vos RC, Moco S, Lommen A, Keurentjes JJ, Bino RJ, Hall RD (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protoc **2**: 778–791

Voytas DF (2013) Plant Genome Engineering with Sequence-Specific Nucleases. Annu Rev Plant Biol **64**: 327–350

Wale SJ (Stuart J., Cattlin ND (2008) Diseases, pests and disorders of potatoes : a colour handbook. Manson Publishing

Wang M, Allefs S, van den Berg RG, Vleeshouwers VGAA, van der Vossen EAG, Vosman B (2008) Allele mining in Solanum: conserved homologues of Rpi-blb1 are identified in Solanum stoloniferum. Theor Appl Genet **116**: 933–943

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **10**: 57–63

Ward JL, Forcat S, Beckmann M, Bennett M, Miller SJ, Baker JM, Hawkins ND, Vermeer CP, Lu C, Lin W, et al (2010) The metabolic transition during disease following infection of Arabidopsis thaliana by Pseudomonas syringae pv. tomato. Plant J **63**: 443–457

Wiesel L, Davis JL, Milne L, Redondo Fernandez V, Herold MB, Middlefell Williams J, Morris J, Hedley PE, Harrower B, Newton AC, et al (2015) A transcriptional reference map of defence hormone responses in potato. Sci Rep **5**: 15229

Witek K, Jupe F, Witek AI, Baker D, Clark MD, Jones JDG (2016) Accelerated cloning of a potato late blight–resistance gene using RenSeq and SMRT sequencing. Nat Biotechnol **34**: 656–660

Woo JW, Kim J, Kwon S Il, Corvalán C, Cho SW, Kim H, Kim S-G, Kim S-T, Choe S, Kim J-S (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nat Biotechnol **33**: 1162–1164

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J, et al (2011) Genome sequence and analysis of the tuber crop potato. Nature **475**: 189–195

Yogendra KN, Dhokane D, Kushalappa AC, Sarmiento F, Rodriguez E, Mosquera T (2017a) StWRKY8 transcription factor regulates benzylisoquinoline alkaloid pathway in potato conferring resistance to late blight. Plant Sci **256**: 208–216

Yogendra KN, Dhokane D, Kushalappa AC, Sarmiento F, Rodriguez E, Mosquera T (2017b) StWRKY8 transcription factor regulates benzylisoquinoline alkaloid pathway in potato conferring resistance to late blight. Plant Sci **256**: 208–216

Yogendra KN, Kumar A, Sarkar K, Li Y, Pushpa D, Mosa KA, Duggavathi R, Kushalappa AC (2015a) Transcription factor StWRKY1 regulates phenylpropanoid

metabolites conferring late blight resistance in potato. J Exp Bot **66**: 7377–89

Yogendra KN, Kushalappa AC (2016) Integrated transcriptomics and metabolomics reveal induction of hierarchies of resistance genes in potato against late blight. Funct Plant Biol **15**: 497–7828

Yogendra KN, Kushalappa AC, Sarmiento F, Rodriguez E, Mosquera T (2015b) Metabolomics deciphers quantitative resistance mechanisms in diploid potato clones against late blight. Funct Plant Biol **42**: 284–298

Yogendra KN, Pushpa D, Mosa KA, Kushalappa AC, Murphy A, Mosquera T (2014) Quantitative resistance in potato leaves to late blight associated with induced hydroxycinnamic acid amides. Funct Integr Genomics **14**: 285–298

Yogendra KN, Sarkar K, Kage U, Kushalappa AC (2017c) Potato NAC43 and MYB8 Mediated Transcriptional Regulation of Secondary Cell Wall Biosynthesis to Contain Phytophthora infestans Infection. Plant Mol Biol Report. doi: 10.1007/s11105-017-1043-1

Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, et al (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. Elife **2**: e00731

Zhang R, Marshall D, Bryan GJ, Hornyik C (2013) Identification and Characterization of miRNA Transcriptome in Potato by High-Throughput Sequencing. PLoS One. doi: 10.1371/journal.pone.0057233

Appendix

Appendix I

Resistance related induced metabolites identified in Chapter III

	Metabolites	ID	Identification method	Molecular formula	Fold Change (RRI)	AME (ppm)
Terpenoids						
	Erythroxanthin sulfate	LMPR01070147	LIPIDMAPS search	C40H54O7S	26.8	3.3955692
	(-)-Solenolide A	LMPR01042300003	LIPIDMAPS search	C28H39ClO9	12.9	-0.7230398
	Cymarin	C08859	KEGG search	C30H44O9	12.08	-0.7308738
	Cerberin	C19984	KEGG search	C32H48O9	11.64	1.91196978
	10-Hydroxyloganin	C11659	KEGG search	C17H26O11	9.43	1.88280472
	Oleanoic acid 3-O-glucuronide	C08964	KEGG search	C36H56O9	9.18	0.62735118
	Kanokoside D	C17431	KEGG search	C27H44O16	5.51	-0.0178097
	Vernoflexuoside	C09579	KEGG search	C21H28O8	4.55	2.6593218
	Officinalisnin	C08904	KEGG search	C45H76O19	4.18	-2.072566
	Dehydrovomifoliol	C02533	KEGG search	C13H18O3	3.89	0.63088561
	7beta,12alpha-Dihydroxykaurenolide	C09081	KEGG search	C20H28O4	3.83	-1.9106847

	Patrinoside	C11640	KEGG search	C21H34O11	3.7	3.83366256
	Secologanin	C01852	KEGG search	C17H24O10	3.21	3.73491775
Flavonoids						
	8-Prenylnaringenin	C18023	KEGG search	C20H20O5	9.01	1.67018718
	Vitexin 2''-O-beta-D-glucoside	C04024	KEGG search	C27H30O15	4.89	0.84464158
	Quercetin 3-sophoroside	C12667	KEGG search	C27H30O17	4.85	1.63978921
	Anhydroicaritin 3-rhamnosyl-(1->2)-rhamnoside	LMPK12112007	LIPIDMAPS search	C33H40O14	4.44	4.85409648
	Karanjin	LMPK12111542	LIPIDMAPS search	C18H12O4	3.46	4.08840871
	Kaempferol 3-O-glucoside	C12249	KEGG search	C21H20O11	3.33	1.37197341
	Lupinisoflavone J	LMPK12050282	LIPIDMAPS search	C25H26O7	3.04	-2.0823088
Alkaloids						
	Alangicine	C09327	KEGG search	C28H36N2O5	7.34	2.29522666
	Serratine	C09901	KEGG search	C16H25NO3	4.64	3.80118436
	Solanine	C10820	KEGG search	C45H73NO15	2.66	1.78143689
Phenylpropanoid						
	Feruloylagmatine	C01670	KEGG search	C15H22N4O3	12.53	3.43489408
	Robustaol A	C09968	KEGG search	C25H30O9	12.18	-2.372248
	Dihydroconiferyl alcohol	C10448	KEGG search	C10H14O3	3.39	-1.0199567

	Sinapoyl malate	C02887	KEGG search	C15H16O9	3.01	3.30430788
	Steganacin	C10886	KEGG search	C24H24O9	3.79	-2.0894989
Fatty acids and lipids						
	N-heptanoyl-homoserine lactone	LMFA08030006	LIPIDMAPS search	C11H19NO3	7.34	0.997979
	PI(17:2(9Z,12Z)/0:0)	LMGP06050015	LIPIDMAPS search	C26H47O12P	4.33	-0.2011105
	(9Z)-(7S,8S)-Dihydroxyoctadecenoic acid	C07355	KEGG search	C18H34O4	3.84	2.05274997
	6,9,12,15-octadecatetraenoic acid	LMFA01030169	LIPIDMAPS search	C18H28O2	3.79	4.17039048
	2-pentadecenoic acid	LMFA01030052	LIPIDMAPS search	C15H28O2	3.2	3.23032858

Appendix II

Resistance related induced genes identified in Chapter III

PGSC_ID	RRI Log_2_FC
PGSC0003DMT400059095	1.501
PGSC0003DMT400054822	1.503
PGSC0003DMT400000931	1.504
PGSC0003DMT400049821	1.505
PGSC0003DMT400030538	1.508
PGSC0003DMT400032575	1.509
PGSC0003DMT400001303	1.514
PGSC0003DMT400009952	1.517
PGSC0003DMT400056509	1.525
PGSC0003DMT400001716	1.527
PGSC0003DMT400083995	1.527
PGSC0003DMT400031242	1.528
PGSC0003DMT400039410	1.529
PGSC0003DMT400056343	1.531
PGSC0003DMT400024114	1.533
PGSC0003DMT400064016	1.536
PGSC0003DMT400008513	1.536
PGSC0003DMT400016846	1.539
PGSC0003DMT400028822	1.541
PGSC0003DMT400079981	1.542
PGSC0003DMT400064344	1.547
PGSC0003DMT400077205	1.547
PGSC0003DMT400078735	1.550
PGSC0003DMT400040781	1.553
PGSC0003DMT400001192	1.553
PGSC0003DMT400050221	1.554
PGSC0003DMT400063267	1.555
PGSC0003DMT400075601	1.558
PGSC0003DMT400037398	1.562
PGSC0003DMT400004851	1.564
PGSC0003DMT400078202	1.566
PGSC0003DMT400071281	1.570
PGSC0003DMT400031827	1.574
PGSC0003DMT400039281	1.577
PGSC0003DMT400041350	1.580
PGSC0003DMT400044389	1.585

PGSC0003DMT400010420	1.594
PGSC0003DMT400081375	1.596
PGSC0003DMT400083985	1.597
PGSC0003DMT400009373	1.599
PGSC0003DMT400049634	1.600
PGSC0003DMT400057449	1.601
PGSC0003DMT400009489	1.604
PGSC0003DMT400056563	1.605
PGSC0003DMT400000646	1.606
PGSC0003DMT400012843	1.609
PGSC0003DMT400046253	1.619
PGSC0003DMT400092392	1.621
PGSC0003DMT400054945	1.628
PGSC0003DMT400050017	1.630
PGSC0003DMT400056852	1.637
PGSC0003DMT400084210	1.640
PGSC0003DMT400066392	1.646
PGSC0003DMT400041756	1.647
PGSC0003DMT400023368	1.651
PGSC0003DMT400067264	1.656
PGSC0003DMT400023917	1.657
PGSC0003DMT400066905	1.659
PGSC0003DMT400061587	1.660
PGSC0003DMT400072411	1.664
PGSC0003DMT400077286	1.678
PGSC0003DMT400035984	1.685
PGSC0003DMT400076378	1.686
PGSC0003DMT400055978	1.691
PGSC0003DMT400003995	1.692
PGSC0003DMT400036715	1.698
PGSC0003DMT400023029	1.711
PGSC0003DMT400051127	1.712
PGSC0003DMT400025919	1.715
PGSC0003DMT400057781	1.720
PGSC0003DMT400000536	1.721
PGSC0003DMT400056773	1.723
PGSC0003DMT400073794	1.726
PGSC0003DMT400038132	1.730
PGSC0003DMT400057305	1.733
PGSC0003DMT400011342	1.733
PGSC0003DMT400055473	1.738
PGSC0003DMT400007545	1.740
PGSC0003DMT400007870	1.740
PGSC0003DMT400067603	1.740

PGSC0003DMT400078304	1.741
PGSC0003DMT400047494	1.742
PGSC0003DMT400037401	1.748
PGSC0003DMT400006390	1.767
PGSC0003DMT400054387	1.769
PGSC0003DMT400024193	1.769
PGSC0003DMT400018869	1.771
PGSC0003DMT400024782	1.771
PGSC0003DMT400045466	1.777
PGSC0003DMT400067426	1.779
PGSC0003DMT400028763	1.780
PGSC0003DMT400039659	1.781
PGSC0003DMT400046634	1.793
PGSC0003DMT400074781	1.794
PGSC0003DMT400041762	1.795
PGSC0003DMT400051023	1.796
PGSC0003DMT400009196	1.807
PGSC0003DMT400031184	1.807
PGSC0003DMT400032781	1.807
PGSC0003DMT400054780	1.820
PGSC0003DMT400023629	1.830
PGSC0003DMT400052206	1.830
PGSC0003DMT400053349	1.841
PGSC0003DMT400041591	1.846
PGSC0003DMT400059654	1.847
PGSC0003DMT400083137	1.851
PGSC0003DMT400021919	1.853
PGSC0003DMT400031901	1.855
PGSC0003DMT400041985	1.856
PGSC0003DMT400052365	1.861
PGSC0003DMT400012950	1.871
PGSC0003DMT400047171	1.873
PGSC0003DMT400011422	1.874
PGSC0003DMT400069776	1.874
PGSC0003DMT400008785	1.874
PGSC0003DMT400001238	1.882
PGSC0003DMT400043125	1.890
PGSC0003DMT400028917	1.890
PGSC0003DMT400040298	1.896
PGSC0003DMT400043522	1.906
PGSC0003DMT400013297	1.915
PGSC0003DMT400018331	1.926
PGSC0003DMT400034871	1.926
PGSC0003DMT400067687	1.939

PGSC0003DMT400003859	1.939
PGSC0003DMT400001663	1.941
PGSC0003DMT400079427	1.942
PGSC0003DMT400023208	1.946
PGSC0003DMT400047737	1.947
PGSC0003DMT400010342	1.955
PGSC0003DMT400078712	1.955
PGSC0003DMT400069969	1.956
PGSC0003DMT400041704	1.960
PGSC0003DMT400087482	1.964
PGSC0003DMT400047579	1.966
PGSC0003DMT400032510	1.968
PGSC0003DMT400033107	1.975
PGSC0003DMT400018673	1.986
PGSC0003DMT400057328	1.988
PGSC0003DMT400068437	1.996
PGSC0003DMT400033366	2.001
PGSC0003DMT400026840	2.004
PGSC0003DMT400068302	2.011
PGSC0003DMT400046670	2.012
PGSC0003DMT400041065	2.016
PGSC0003DMT400041944	2.022
PGSC0003DMT400019373	2.022
PGSC0003DMT400038414	2.024
PGSC0003DMT400078290	2.024
PGSC0003DMT400065094	2.025
PGSC0003DMT400029590	2.028
PGSC0003DMT400068226	2.033
PGSC0003DMT400077358	2.037
PGSC0003DMT400068648	2.038
PGSC0003DMT400085139	2.039
PGSC0003DMT400048875	2.041
PGSC0003DMT400029056	2.043
PGSC0003DMT400064211	2.043
PGSC0003DMT400006713	2.050
PGSC0003DMT400055296	2.055
PGSC0003DMT400048331	2.060
PGSC0003DMT400044065	2.066
PGSC0003DMT400063316	2.074
PGSC0003DMT400068062	2.081
PGSC0003DMT400008504	2.085
PGSC0003DMT400056932	2.088
PGSC0003DMT400057777	2.095
PGSC0003DMT400002162	2.097

PGSC0003DMT400064988	2.098
PGSC0003DMT400077751	2.103
PGSC0003DMT400005136	2.103
PGSC0003DMT400001512	2.104
PGSC0003DMT400009479	2.106
PGSC0003DMT400058309	2.113
PGSC0003DMT400063438	2.115
PGSC0003DMT400023921	2.117
PGSC0003DMT400067234	2.126
PGSC0003DMT400028172	2.129
PGSC0003DMT400048627	2.131
PGSC0003DMT400040010	2.140
PGSC0003DMT400017981	2.151
PGSC0003DMT400041812	2.152
PGSC0003DMT400072912	2.152
PGSC0003DMT400015480	2.153
PGSC0003DMT400075401	2.154
PGSC0003DMT400016521	2.155
PGSC0003DMT400062861	2.164
PGSC0003DMT400023883	2.167
PGSC0003DMT400061392	2.168
PGSC0003DMT400008271	2.169
PGSC0003DMT400035841	2.170
PGSC0003DMT400007545	2.178
PGSC0003DMT400063703	2.182
PGSC0003DMT400015906	2.182
PGSC0003DMT400002565	2.184
PGSC0003DMT400005613	2.186
PGSC0003DMT400000067	2.189
PGSC0003DMT400007048	2.190
PGSC0003DMT400001620	2.193
PGSC0003DMT400090661	2.194
PGSC0003DMT400029874	2.215
PGSC0003DMT400046778	2.215
PGSC0003DMT400079604	2.217
PGSC0003DMT400081229	2.218
PGSC0003DMT400060514	2.226
PGSC0003DMT400004597	2.227
PGSC0003DMT400078518	2.234
PGSC0003DMT400070037	2.238
PGSC0003DMT400076979	2.243
PGSC0003DMT400037425	2.246
PGSC0003DMT400023897	2.246
PGSC0003DMT400032758	2.252

PGSC0003DMT400072700	2.258
PGSC0003DMT400005637	2.267
PGSC0003DMT400078294	2.270
PGSC0003DMT400076297	2.272
PGSC0003DMT400046238	2.274
PGSC0003DMT400004618	2.279
PGSC0003DMT400079121	2.284
PGSC0003DMT400000780	2.288
PGSC0003DMT400045324	2.291
PGSC0003DMT400060883	2.312
PGSC0003DMT400002555	2.312
PGSC0003DMT400041847	2.323
PGSC0003DMT400013021	2.333
PGSC0003DMT400021489	2.333
PGSC0003DMT400049235	2.335
PGSC0003DMT400052184	2.337
PGSC0003DMT400037826	2.339
PGSC0003DMT400040147	2.346
PGSC0003DMT400010297	2.347
PGSC0003DMT400046711	2.353
PGSC0003DMT400027316	2.355
PGSC0003DMT400042984	2.359
PGSC0003DMT400026141	2.362
PGSC0003DMT400081182	2.368
PGSC0003DMT400046950	2.386
PGSC0003DMT400027659	2.388
PGSC0003DMT400035284	2.389
PGSC0003DMT400025631	2.391
PGSC0003DMT400018015	2.402
PGSC0003DMT400024113	2.404
PGSC0003DMT400084231	2.421
PGSC0003DMT400073300	2.424
PGSC0003DMT400049124	2.430
PGSC0003DMT400054207	2.442
PGSC0003DMT400026198	2.443
PGSC0003DMT400044107	2.444
PGSC0003DMT400020638	2.447
PGSC0003DMT400058632	2.459
PGSC0003DMT400037403	2.461
PGSC0003DMT400077913	2.483
PGSC0003DMT400062354	2.487
PGSC0003DMT400017899	2.497
PGSC0003DMT400063313	2.501
PGSC0003DMT400059197	2.504

PGSC0003DMT400067667	2.507
PGSC0003DMT400046207	2.507
PGSC0003DMT400010340	2.513
PGSC0003DMT400079196	2.520
PGSC0003DMT400000884	2.521
PGSC0003DMT400082139	2.533
PGSC0003DMT400007095	2.535
PGSC0003DMT400005798	2.538
PGSC0003DMT400078688	2.539
PGSC0003DMT400060849	2.544
PGSC0003DMT400081914	2.548
PGSC0003DMT400055211	2.553
PGSC0003DMT400031867	2.555
PGSC0003DMT400058417	2.558
PGSC0003DMT400000507	2.558
PGSC0003DMT400079099	2.566
PGSC0003DMT400013398	2.578
PGSC0003DMT400029242	2.579
PGSC0003DMT400081445	2.593
PGSC0003DMT400076435	2.601
PGSC0003DMT400054502	2.614
PGSC0003DMT400012086	2.614
PGSC0003DMT400032497	2.622
PGSC0003DMT400036505	2.624
PGSC0003DMT400075600	2.627
PGSC0003DMT400086945	2.633
PGSC0003DMT400018016	2.644
PGSC0003DMT400060191	2.662
PGSC0003DMT400060477	2.666
PGSC0003DMT400076481	2.668
PGSC0003DMT400080128	2.683
PGSC0003DMT400040837	2.695
PGSC0003DMT400019290	2.702
PGSC0003DMT400075612	2.715
PGSC0003DMT400013376	2.726
PGSC0003DMT400046260	2.748
PGSC0003DMT400069660	2.752
PGSC0003DMT400083779	2.758
PGSC0003DMT400075056	2.760
PGSC0003DMT400047645	2.763
PGSC0003DMT400026700	2.766
PGSC0003DMT400068438	2.768
PGSC0003DMT400077055	2.778
PGSC0003DMT400055930	2.784

PGSC0003DMT400074356	2.788
PGSC0003DMT400056685	2.788
PGSC0003DMT400026261	2.794
PGSC0003DMT400061480	2.799
PGSC0003DMT400018729	2.800
PGSC0003DMT400000423	2.801
PGSC0003DMT400053338	2.805
PGSC0003DMT400053483	2.806
PGSC0003DMT400079966	2.807
PGSC0003DMT400065260	2.822
PGSC0003DMT400003869	2.828
PGSC0003DMT400044692	2.830
PGSC0003DMT400051027	2.837
PGSC0003DMT400048873	2.848
PGSC0003DMT400075057	2.850
PGSC0003DMT400051168	2.861
PGSC0003DMT400039840	2.868
PGSC0003DMT400079611	2.882
PGSC0003DMT400033395	2.886
PGSC0003DMT400012976	2.888
PGSC0003DMT400033585	2.897
PGSC0003DMT400043856	2.899
PGSC0003DMT400059824	2.913
PGSC0003DMT400093652	2.931
PGSC0003DMT400074766	2.931
PGSC0003DMT400063921	2.944
PGSC0003DMT400075163	2.945
PGSC0003DMT400019058	2.945
PGSC0003DMT400062986	2.953
PGSC0003DMT400054816	2.956
PGSC0003DMT400004061	2.959
PGSC0003DMT400079265	2.960
PGSC0003DMT400037925	2.964
PGSC0003DMT400034336	2.967
PGSC0003DMT400072587	2.968
PGSC0003DMT400009328	2.971
PGSC0003DMT400065402	2.972
PGSC0003DMT400039670	2.973
PGSC0003DMT400008634	2.977
PGSC0003DMT400003877	2.986
PGSC0003DMT400067663	2.994
PGSC0003DMT400040485	3.004
PGSC0003DMT400005177	3.024
PGSC0003DMT400078104	3.024

PGSC0003DMT400001785	3.026
PGSC0003DMT400010886	3.030
PGSC0003DMT400033302	3.073
PGSC0003DMT400020647	3.084
PGSC0003DMT400008432	3.131
PGSC0003DMT400035365	3.143
PGSC0003DMT400061140	3.152
PGSC0003DMT400067258	3.153
PGSC0003DMT400044514	3.178
PGSC0003DMT400009308	3.184
PGSC0003DMT400013130	3.201
PGSC0003DMT400022255	3.248
PGSC0003DMT400026138	3.248
PGSC0003DMT400073709	3.249
PGSC0003DMT400088620	3.261
PGSC0003DMT400064290	3.267
PGSC0003DMT400009296	3.276
PGSC0003DMT400009982	3.277
PGSC0003DMT400064225	3.286
PGSC0003DMT400028619	3.287
PGSC0003DMT400002771	3.301
PGSC0003DMT400036934	3.303
PGSC0003DMT400080759	3.310
PGSC0003DMT400042837	3.316
PGSC0003DMT400070963	3.322
PGSC0003DMT400010463	3.341
PGSC0003DMT400067990	3.342
PGSC0003DMT400080788	3.348
PGSC0003DMT400004111	3.366
PGSC0003DMT400018673	3.382
PGSC0003DMT400078152	3.420
PGSC0003DMT400044688	3.453
PGSC0003DMT400043508	3.454
PGSC0003DMT400069792	3.457
PGSC0003DMT400062372	3.470
PGSC0003DMT400071337	3.471
PGSC0003DMT400040186	3.480
PGSC0003DMT400045755	3.485
PGSC0003DMT400063292	3.486
PGSC0003DMT400021541	3.494
PGSC0003DMT400018441	3.508
PGSC0003DMT400034271	3.512
PGSC0003DMT400040307	3.562
PGSC0003DMT400043113	3.565

PGSC0003DMT400053967	3.567
PGSC0003DMT400032619	3.568
PGSC0003DMT400036399	3.569
PGSC0003DMT400064456	3.571
PGSC0003DMT400058159	3.578
PGSC0003DMT400065644	3.614
PGSC0003DMT400090515	3.624
PGSC0003DMT400005614	3.633
PGSC0003DMT400026784	3.639
PGSC0003DMT400066574	3.646
PGSC0003DMT400051886	3.656
PGSC0003DMT400009831	3.657
PGSC0003DMT400039815	3.667
PGSC0003DMT400011205	3.676
PGSC0003DMT400033347	3.685
PGSC0003DMT400007615	3.697
PGSC0003DMT400011027	3.711
PGSC0003DMT400058117	3.716
PGSC0003DMT400059566	3.735
PGSC0003DMT400043311	3.747
PGSC0003DMT400059487	3.753
PGSC0003DMT400033866	3.761
PGSC0003DMT400023470	3.773
PGSC0003DMT400073281	3.776
PGSC0003DMT400011810	3.778
PGSC0003DMT400004910	3.779
PGSC0003DMT400059673	3.785
PGSC0003DMT400061220	3.817
PGSC0003DMT400035267	3.824
PGSC0003DMT400036669	3.830
PGSC0003DMT400070196	3.833
PGSC0003DMT400069232	3.848
PGSC0003DMT400078489	3.881
PGSC0003DMT400062145	3.893
PGSC0003DMT400003773	3.902
PGSC0003DMT400050821	3.907
PGSC0003DMT400024344	3.919
PGSC0003DMT400060865	3.921
PGSC0003DMT400041984	3.928
PGSC0003DMT400056122	3.939
PGSC0003DMT400059283	3.962
PGSC0003DMT400015899	3.992
PGSC0003DMT400008902	4.002
PGSC0003DMT400015066	4.029

PGSC0003DMT400063790	4.039
PGSC0003DMT400001475	4.048
PGSC0003DMT400051635	4.059
PGSC0003DMT400044243	4.077
PGSC0003DMT400028369	4.109
PGSC0003DMT400030337	4.142
PGSC0003DMT400055916	4.159
PGSC0003DMT400012347	4.195
PGSC0003DMT400036420	4.199
PGSC0003DMT400076200	4.224
PGSC0003DMT400019129	4.226
PGSC0003DMT400026840	4.251
PGSC0003DMT400066570	4.253
PGSC0003DMT400028227	4.254
PGSC0003DMT400040521	4.266
PGSC0003DMT400067782	4.298
PGSC0003DMT400064185	4.309
PGSC0003DMT400007865	4.314
PGSC0003DMT400064346	4.315
PGSC0003DMT400082479	4.319
PGSC0003DMT400013501	4.353
PGSC0003DMT400010110	4.358
PGSC0003DMT400003360	4.359
PGSC0003DMT400069635	4.461
PGSC0003DMT400016494	4.477
PGSC0003DMT400069961	4.483
PGSC0003DMT400048941	4.489
PGSC0003DMT400097075	4.511
PGSC0003DMT400065792	4.532
PGSC0003DMT400049371	4.628
PGSC0003DMT400040116	4.670
PGSC0003DMT400065346	4.670
PGSC0003DMT400066209	4.692
PGSC0003DMT400012346	4.695
PGSC0003DMT400013972	4.725
PGSC0003DMT400035395	4.736
PGSC0003DMT400072044	4.737
PGSC0003DMT400067343	4.752
PGSC0003DMT400061283	4.774
PGSC0003DMT400074105	4.879
PGSC0003DMT400080885	4.903
PGSC0003DMT400002504	4.907
PGSC0003DMT400016038	4.907
PGSC0003DMT400007341	4.930

PGSC0003DMT400050183	4.940
PGSC0003DMT400037720	5.035
PGSC0003DMT400028546	5.059
PGSC0003DMT400016636	5.116
PGSC0003DMT400066925	5.160
PGSC0003DMT400013769	5.163
PGSC0003DMT400048344	5.180
PGSC0003DMT400073253	5.186
PGSC0003DMT400026627	5.219
PGSC0003DMT400079071	5.259
PGSC0003DMT400049613	5.264
PGSC0003DMT400083083	5.299
PGSC0003DMT400073715	5.317
PGSC0003DMT400054145	5.322
PGSC0003DMT400000538	5.360
PGSC0003DMT400040774	5.433
PGSC0003DMT400065043	5.443
PGSC0003DMT400046567	5.459
PGSC0003DMT400052065	5.555
PGSC0003DMT400037083	5.561
PGSC0003DMT400056432	5.605
PGSC0003DMT400080508	5.656
PGSC0003DMT400073334	5.656
PGSC0003DMT400017417	5.672
PGSC0003DMT400028797	5.716
PGSC0003DMT400048379	5.848
PGSC0003DMT400049664	5.887
PGSC0003DMT400022827	6.146
PGSC0003DMT400039134	6.199
PGSC0003DMT400036828	6.209
PGSC0003DMT400055393	6.264
PGSC0003DMT400067535	6.284
PGSC0003DMT400024250	6.363
PGSC0003DMT400078178	6.435
PGSC0003DMT400083871	6.444
PGSC0003DMT400001824	6.492
PGSC0003DMT400085905	6.627
PGSC0003DMT400007275	6.655
PGSC0003DMT400043146	6.662
PGSC0003DMT400069259	6.665
PGSC0003DMT400066572	6.728
PGSC0003DMT400005546	6.759
PGSC0003DMT400069898	6.931
PGSC0003DMT400023280	7.219

PGSC0003DMT400014567	7.230
PGSC0003DMT400032694	7.322
PGSC0003DMT400048241	7.324
PGSC0003DMT400039061	7.509
PGSC0003DMT400023141	7.545
PGSC0003DMT400049361	7.582
PGSC0003DMT400044959	7.599
PGSC0003DMT400043662	7.784
PGSC0003DMT400079684	7.807
PGSC0003DMT400039067	7.813
PGSC0003DMT400020468	7.882
PGSC0003DMT400066947	7.943
PGSC0003DMT400039492	8.109
PGSC0003DMT400033302	8.648
PGSC0003DMT400023792	8.659
PGSC0003DMT400044207	8.680
PGSC0003DMT400075503	8.719
PGSC0003DMT400005250	8.738
PGSC0003DMT400044559	8.903
PGSC0003DMT400050908	9.189
PGSC0003DMT400037771	9.278
PGSC0003DMT400083089	9.434
PGSC0003DMT400044229	9.475
PGSC0003DMT400041449	9.643
PGSC0003DMT400017879	10.127
PGSC0003DMT400065943	10.195
PGSC0003DMT400037143	10.270
PGSC0003DMT400042707	10.789
PGSC0003DMT400031195	11.200
PGSC0003DMT400027947	11.644
PGSC0003DMT400074744	11.648
PGSC0003DMT400026448	11.690
PGSC0003DMT400079860	11.704
PGSC0003DMT400026291	11.845
PGSC0003DMT400042840	11.873
PGSC0003DMT400031899	11.884
PGSC0003DMT400014960	12.022
PGSC0003DMT400078265	12.167
PGSC0003DMT400002959	12.279
PGSC0003DMT400034097	12.300
PGSC0003DMT400047219	12.399
PGSC0003DMT400013091	12.411
PGSC0003DMT400056154	12.592
PGSC0003DMT400046038	12.659

PGSC0003DMT400044065	12.750
PGSC0003DMT400041381	13.397
PGSC0003DMT400027951	13.748
PGSC0003DMT400012344	14.023
PGSC0003DMT400010004	14.392
PGSC0003DMT400002013	14.963

- The details of PGSC IDs can be found at the potato genome consortium website http://solanaceae.plantbiology.msu.edu/
- RRI= Resistant related induced, FC= Fold change

Appendix III

Primary TFs identified in Chapter III

Description	PGSC_ID	Log_2_FC (RRI)	Uniprot ID
NAC domain-containing protein 2	PGSC0003DMT400023917	1.657	Q39013
WRKY transcription factor 23	PGSC0003DMT400023368	1.651	O22900
E2F transcription factor-like E2FE	PGSC0003DMT400062861	2.164	Q8LSZ4
Ethylene-responsive factor RAP2-3	PGSC0003DMT400043311	3.747	P42736
Heat stress transcription factor B-3	PGSC0003DMT400070196	3.833	O22230
bZIP transcription factor TGA10	PGSC0003DMT400060865	3.921	Q52MZ2
NAC transcription factor 29	PGSC0003DMT400007341	4.930	O49255
Agamous-like MADS-box protein AGL15	PGSC0003DMT400017417	5.672	Q39295
LOB domain-containing protein 15	PGSC0003DMT400078265	12.167	Q8L5T5
Transcription factor MYB61	PGSC0003DMT400040774	5.433	Q8VZQ2
Probable WRKY transcription factor 51	PGSC0003DMT400080128	2.683	Q93WU9
NAC transcription factor 56	PGSC0003DMT400049664	5.887	Q9LD44
Transcription factor bHLH66	PGSC0003DMT400028917	1.890	Q9ZUG9
Transcription factor HHO2	PGSC0003DMT400002162	2.097	Q8VZS3

Appendix IV

Secondary TFs identified in Chapter III

Description	Log2 FC (RRI)	PGSC_Transcript_ID	Uniprot ID
Flowering-promoting factor 1-like protein 3	3.78	PGSC0003DMT400011810	Q0E1D7
LBD 15	12.17	PGSC0003DMT400078265	Q8L5T5
NAC 73	2.22	PGSC0003DMT400081229	O49459
NAC 104	2.67	PGSC0003DMT400076481	Q8GWK6
MADS-box 27	1.77	PGSC0003DMT400024193	Q6EP49
bZIP 11	4.72	PGSC0003DMT400013972	O65683
TCP20	2.19	PGSC0003DMT400090661	Q9LSD5
MYB102	12.40	PGSC0003DMT400047219	Q9LDR8
BEL1-like 9	2.33	PGSC0003DMT400049235	Q9LZM8
AT-hook 15	3.46	PGSC0003DMT400069792	Q9M2S3
WRKY 40	2.95	PGSC0003DMT400019058	Q9SAH7
LBD 38	6.26	PGSC0003DMT400055393	Q9SN23
MYB48	3.42	PGSC0003DMT400078152	Q9LX82
bZIP TGA10	3.92	PGSC0003DMT400060865	Q52MZ2
LBD 11	4.67	PGSC0003DMT400065346	Q9SK08
WRKY 51	2.68	PGSC0003DMT400080128	Q93WU9
NAC 72	2.97	PGSC0003DMT400039670	Q93VY3
Transcription repressor KAN1	2.15	PGSC0003DMT400017981	Q93WJ9
NAC104	2.67	PGSC0003DMT400076481	Q8GWK6
MADS-box 27	1.77	PGSC0003DMT400024193	Q6EP49

Appendix V

Genes selected for qPCR with the primers list used for qPCR analysis

qPCR Table				
Gene	**Genbank ID**	**PGSC ID**	**Forward Primer**	**Reverse Primer**
Peroxidase 17	XM_006353151.2	PGSC0003DMT400041812	AGAGATGCTGTTGTTCTGAGTGG	TGCATTTGATCTTGGACTTGGC
Alcohol dehydrogenase 3	NM_001288151.1	PGSC0003DMT400079071	TGCTCCTCCACAGAAAATGGA	ACACTCTCCACAATCCCTGC
WIN1	NM_001288699.1	PGSC0003DMT400050017	GCCTGGCGGAGTAAGTATGG	ATCCACGATTCTCACCGTCG
STH-2	XM_006340827.2	PGSC0003DMT400003773	TGCCCCAACAAGGCTATTC	GGACCACCTTCAACAAAGTTC
Neoxanthin synthase	NM_001318672.1	PGSC0003DMT400015066	TTCAGGTCGGGCTCAATTCG	TTGGCCACATGGAAAGTGGT
NAC72	XM_006344202.2	PGSC0003DMT400039670	GGGAATTGCCTGAAAAAGCTGAA	ATCTGTGCCTGTGGCTTTCC
MYB48	XM_015310207.1	PGSC0003DMT400078152	AGAGTTGCAGATTACGATGGGTT	AGCTATTTTCGACCATCTATTCCCA
4CL like 9	XM_006351420.2	PGSC0003DMT400039281	GGCAGAGCGGTTCAAGAGCAGG	CCTCAGGGCCATTCATCCCTGTT

Appendix VI

Primers used in Chapter IV

	Forward Primer	**Reverse Primer**
StACT1	CCCTACTTCTTAATTTCTTTGTCC	ACGACTCATTTAATTTGTGTGAC
StACT2	CGAGCTTAGCCATTCTTCTCTTC	GCAAATGAATTTTCATATCATCACATAG
StFT	CCACACTCACAAGTATACCATT	TGATCTCAATGGGATGGAGATA
StACT1_qPCR	GGTCATTCCACTAGCAACTTCTTGG	GGGGTTGAATTCATGTCTAAGTCGAA
StACT2_qPCR	AACGCCGCGATTCAATTGGG	TATCGGCAGATGCCTCAACGAA
StFT_qPCR	GCCGTGCCGGTTCGAACAAT	CCACAATAAGCTTCATTTCCTGGC
StEF-1a_qPCR	ATTGGAAACGGATATGCTCCA	TCCTTACCTGAACGCCTGTCA
StACT2_Promoter	GATGTCCATCATAACACGATGATAG	GAACTTTCTGTTTTGACCTTCATGG
StNAC72	ACAACAAATTCGCGTTAAGGTCT	TGCAACGCTTTTCTTCTCTCTC
StNAC72_Promoter	GATGATCGGTGGCTAACTTTATAAA	GAGAGAGAAGAAAAGCGTTGC
Trans_Confirmation	CAAAATTTTTAAAAGTTAACAATGATT	CTCGGTGTATTAAGCTTTTACTATG
NAC_gRNA1	GGATATGGAGAAAATTTTGG	
NAC_gRNA2	CTGCAGGGCCCATTCGGGGA	
NAC_gRNA3	AATTGGCCGACCTTATCTAA	