Genome editing, to replace non-functional genes in the phenylpropanoid pathway, in Russet Burbank potato to enhance resistance against late blight caused by *Phytophthora infestans*

Niranjan Hegde

Department of Plant Science

McGill University, Montreal, Canada

April 2020

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

@ Niranjan Hegde 2020

Dedicated to my beloved parents, grandparents and village farmers

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

AME	Accurate mass error
AUDPC	Area under disease progress curve
BLAST	Basic local alignment search tool
CCoAOMT	Caffeoyl-CoA O-methyltransferase
cDNA	Complimentary deoxyribonucleic acid
DNA	Deoxyribonucleic acid
dpi	Days post inoculation
FAO	Food and agriculture organization
GC-MS	Gas chromatography mass spectrometry
GFP	Green fluorescent protein
HCAAs	Hydroxycinnamic acid amides
InDels	Insertions and Deletions
hpi	Hours post inoculation
kDa	Kilo Dalton
LC-MS	Liquid chromatography mass spectrometry
LC-HRMS	Liquid chromatography high resolution mass spectrometry
m/z ratio	Mass to charge ratio
MS	Mass spectrometry
PCR	Polymerase chain reaction
PR	Pathogenesis related
QTL	Quantitative trait loci
RCBD	Randomized complete block design
RM	Mock inoculated resistant genotype
RNA	Ribonucleic Acid
RP	Pathogen inoculated resistant genotype
RR	Resistance related
RRC	Resistance related constitutive metabolite
RRI	Resistance related induced metabolite
SM	Mock inoculated susceptible genotype

SNP	Single nucleotide polymorphism
SP	Pathogen inoculated susceptible genotype
SSR	Simple Sequence Repeat

ABSTRACT

Late blight, caused by an oomycete *Phytophthora infestans*, is the most devastating disease of potato worldwide and remains as the major threat to potato production. Resistance to late blight is highly complex and considered to be qualitative with a hypersensitive response or quantitative with reduced susceptibility. The qualitative response is mainly due to receptor R genes, whereas the quantitative resistance is primarily due to constitutive or induced metabolites and proteins. Pyramiding of leucine-rich-repeat (NB-LRR) receptor R genes and fungicide applications are commonly used to manage late blight. However, the receptor R genes are not stable, and the fungicide applications are often not adequate. Pathogen-induced metabolites can impart durable resistance to late blight through active cell wall reinforcement at the site of infection. The resistance-related (RR) metabolites are biosynthesized by a hierarchy of resistance genes, involving regulatory and downstream metabolite biosynthetic genes. This study focused mainly on resistance gene discovery and editing to enhance late blight resistance in a commercial potato cultivar, Russet Burbank. Transcriptomics data on late blight resistant (F06037) and susceptible (Russet Burbank) potato genotypes inoculated with *Phytophthora infestans* were analyzed to discover the Single Nucleotide Polymorphisms (SNPs) and Insertions and Deletions (InDels) involving allelic variations in transcripts. The SNPs and InDels were further classified into different groups based on their ability to code for functional proteins. The downstream genes involved in the biosynthesis of RR metabolites in various metabolic pathways were separated, and the SNPs and InDels in the Russet Burbank genotype were identified. In the second study, one of these genes (*StCCoAOMT* encoding Caffeoyl-CoA O-methyltransferase) responsible for the biosynthesis of monolignols, which was mutated in Russet Burbank was edited using CRISPR-Cas9. A precise SNP mutation correction of the *StCCoAOMT* in Russet Burbank potato significantly decreased both disease severity and as well the pathogen biomass in edited plants. Metabolic profiling of control and edited plants, inoculated with mock or pathogen, identified significant increases in the accumulation of RR metabolites involved in suberization and lignification of secondary cell walls around the pathogen-infected area in the edited plants. If this gene is mutated in other commercial cultivars, then it can be edited, provided the rest of the hierarchies of genes to biosynthesize the required RR metabolites are functional, to suppress the

pathogen. The study displayed the potential of CRISPR-Cas9 based tool for the improvement disease resistance in potato.

RÉSUMÉ

Le mildiou de la pomme de terre, causé par un oomycète *Phytophthora infestans*, est la maladie la plus dévastatrice de la pomme de terre dans le monde et reste la principale menace pour la production de pomme de terre. La résistance au mildiou est très complexe et considérée comme qualitative avec une réponse hypersensible ou quantitative avec une sensibilité réduite. La réponse qualitative est principalement due aux récepteurs gènes-R, tandis que la résistance quantitative est principalement due aux métabolites et protéines constitutifs ou induits. Le pyramidage de récepteurs gènes-R avec répétition riche en leucine (NB-LRR) et l'application de fongicides sont couramment utilisés pour gérer le mildiou. Cependant, les récepteurs gènes-R ne sont pas stables et les fongicides sont souvent inadéquats. Les métabolites induits par les pathogènes peuvent transmettre une résistance durable au mildiou grâce au renforcement actif de la paroi cellulaire au site d'infection. Les métabolites liés à la résistance (RR) sont biosynthétisés par une hiérarchie de gènes de résistance, impliquant des gènes de biosynthèse des métabolites régulateurs et en aval. Cette étude s'est concentrée principalement sur la découverte et la modification de gènes de résistance afin d'améliorer la résistance au mildiou pour Russet Burbank, un cultivar commercial de pomme de terre. Des données transcriptomiques sur des génotypes de pomme de terre résistantes (F06037) et sensibles (Russet Burbank) au mildiou qui ont été inoculés avec *Phytophthora infestans* ont été analysées pour découvrir les polymorphismes à nucléotide simple (SNP) et les insertions et délétions (InDels) impliquant des variations alléliques dans les transcriptions. Les SNPs et InDels ont été en outre classés en différents groupes en fonction de leur capacité à coder pour les protéines fonctionnelles. Les gènes en aval impliqués dans la biosynthèse des métabolites RR dans différentes voies métaboliques ont été séparés et les SNPs et InDels dans les génotypes de patates Russet Burbank ont été identifiés. Dans la deuxième étude, l'un des gènes (StCCoAOMT codant pour la Caffeoyl-CoA O-méthyltransférase) responsable de la biosynthèse des monolignols, qui a été muté dans les patates Russet Burbank, a été modifié à l'aide de CRISPR-Cas9. Une correction précise des mutations SNP dans les patates Russet Burbank a diminué de manière significative à la fois la gravité de la maladie ainsi que la biomasse des agents pathogènes dans les plantes modifiées. Le profilage métabolique des plantes témoins et

modifiées, inoculées avec des simulacres ou des pathogènes, a identifié des augmentations significatives de l'accumulation de métabolites RR impliqués dans la subérisation et la lignification dans les plantes modifiées. Si ce gène est muté dans d'autres cultivars commerciaux il peut être modifié.

ACKNOWLEDGMENT

Firstly, I wish to express my sincere appreciation to my supervisor, Dr. Ajjamada C. Kushalappa, for his tremendous support, critical guidance, and mentoring me. He convincingly guided and encouraged me to be professional and perform the right thing even when the tasks got tough. Without his persistent help and constant support, the goal of this project would not have been realized and accomplished.

I wish to show my gratitude and heartfelt thanks to Dr. Suha Jabaji and Dr. Elias Georges, for their suggestions, constructive criticism, which helped me in carrying out this research work - I am deeply grateful to members of the advisory committee. Besides my advisor and advisory committee, I would like to thank the comprehensive exam committee: Dr. Philippe Seguin (Chair), Dr. Suha Jabaji, Dr. Danielle Donnelly, and Dr. Saji George, for their insightful comments and encouragement. I also thank them for the hard question, which incented me to widen my research from various perspectives.

My sincere thanks also go to all the faculty and staff of the Plant Science department. Special thanks to Dr. Suha Jabaji, Dr. Jean-Benoit Charron, Dr. Saji George, and their lab members. I would like to recognize the invaluable assistance that all the plant science staff provided during my study: Guy Rimmer, Ian Ritchie, and S. Dernovici, who gave access to the laboratory and research facilities in the greenhouse. Without their precious support, it would not be possible to conduct this research. I sincerely thank Susan Gregus, L. Flood, D. Chan-hum, and Erin Minnett Cid for their much-needed assistance in administrative tasks.

I would like to express my deep sense of thankfulness to Dr. Rajduggavathi, who provided me with an opportunity to use lab equipment for conducting qRT-PCR. I am grateful to Dr. Denis Faubert and his team members from IRCM, Montreal, for the analysis of metabolites. I am also thankful to Serge Dernovici, technician, Institute of Parasitology, to use lab space for microscopy.

I thank my fellow labmates for constantly engaging in scientific discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last four

years. I wish to express my deepest gratitude to Nancy Soni, Sripad Joshi, Russiachand Singh Heikham, Krishna Kumar, Bikram Poudel, Chaithra Nagaraju for continually engaging me in fruitful scientific discussions. Also, I thank my heartfelt thanks to Doddakhalandra Doddamni, Yogendra Kalenahalli, Shailesh Karre, Dhananjay Dhokane, Maria Kyriakidou, Sai Reddy Achakkagari, Irfan Iqbal, F Nazarian-Firouzabadi, who always motivated me by sharing their scientific work experiences and teaching me many scientific aspects. Yogendra Kalenahalli Deepak Nanjappa and Dr. Tomal Dattaroy for significant mentoring throughout my studies and helped me to pursue doctoral studies. I thank Sofie Gautreau for translating abstract to French.

I am indebted to the McGill University and my sincere thanks for providing me a timely financial aid - such as research assistantship, graduate excellence award, the GREAT award for attending the conference, and international student bursary. My appreciation to the Plant Science department for the GREAT award to attend two conferences and deliver oral presentations of my research work - Phytoprotection: a pillar of a sustainable agroecosystem by Quebec society for the protection of plants (SPPQ) and Plant Canada 2019, University of Guelph. I acknowledge the research project funding by the Natural Sciences and Engineering Council of Canada (NSERC) and partly by the Weston-Loblaw, Seeding Food Innovation.

Finally, I would like to thank my family: my parents, grandparents, and sister for supporting me spiritually throughout writing this thesis and my life in general. I wish to acknowledge the support and great love of my family. They kept me going on, and this work would not have been possible without their input.

CONTRIBUTION TO ORIGINAL KNOWLEDGE

Preface

This thesis work is presented in a manuscript-based format. In this research study, I have used the RNA-seq data of potato late blight susceptible (Russet Burbank) and resistance (F06037 and F06025) genotypes inoculated with *Phytophthora infestans*. RNA-seq data were analyzed to identify SNPs/InDel variations among the pathway genes, anticipating that the deleterious mutations within pathway genes and transcripts encoding biosynthetic enzymes would alter the production of a specific set of resistance-related metabolites in late blight susceptible genotype (Russet Burbank). Several important candidate genes with deleterious mutations within the coding region were identified and are validated by Sanger sequencing and gene expression studies. I also present here the functional characterization of one of the candidate genes. The gene *StCCoAOMT* encoding Caffeoyl-CoA O-methyltransferase was knocked out in late blight resistance genotype (F06037) using CRISPR-Cas9 to decipher the resistance role. The nonsense mutation within the transcript variants of the *StCCoAOMT* identified in Russet Burbank was characterized by carrying out the mutation correction using Gemini virus replicon based CRISPR-Cas9 knock-in. The following elements of this study are considered original scholarship and distinct contributions to knowledge:

Contribution to knowledge

- Variant calling was performed for RNA-seq of late blight susceptible (Russet Burbank) and resistance (F06037 and F06025) genotypes inoculated with *Phytophthora infestans.* A similar pipeline presented in the study can be used for other potato and plant species.
- SNPs and InDels were identified in three potato genotypes using several bioinformatics tools. All the tools mentioned can be used elsewhere for comprehensive SNP/InDels identification.
- The RNA-seq based SNP/InDels data will be made available in the NCBI database for further use in the development of any traits in Russet Burbank.

- All the deleterious mutations (SNPs/InDels) within pathogen-induced biosynthetic genes/ transcripts were categorized into biosynthetic pathways. Important candidate genes and transcripts with deleterious mutations can be considered for further studies.
- Achieved an efficient gene knock-in or gene targeting in Russet Burbank potato genotypes by using a geminiviral replicon based CRISPR-Cas9 tool.
- Genome editing tools were developed for precise genome editing in Russet Burbank and other autotetraploid potato genotypes.
- The gene *StCCoSOMT* gene and transcripts with a nonsense mutation in Russet Burbank were studied. The late blight resistance function of *StCCoAOMT* was validated. The *StCCoAOMT* gene, if mutated in other cultivars, can be edited to enhance late blight resistance.

CONTRIBUTION OF AUTHORS

This thesis involves two studies (Chapter III & IV) presented as two manuscripts according to the McGill guidelines for thesis preparation. The research work performed here was designed entirely by me under the guidance of my supervisor Dr. Ajjamada C. Kushalappa. I conducted all the bioinformatic analyses, analyzed data, and wrote manuscripts. I performed all the laboratory and greenhouse experiments. I wrote both documents and thesis under the supervision of Dr. Ajjamada C. Kushalappa.

The III chapter was co-authored by Dadakhalandar Doddamani, who assisted in RNA-seq by providing the space for computational analysis. Chapter IV was co-authored by Shripad Joshi, who assisted in metabolite analysis and Nancy Soni, who helped in gene expression studies. Ajjamada C. Kushalappa supervised the project and edited the manuscript.

CHAPTER I: GENERAL INTRODUCTION

Potato (*Solanum tuberosum* L.) is one of the important food crops in the world in terms of human consumption. The global estimate of potato production in 2012 was 365 million tonnes globally (FAO 2014). The need to feed more than a billion people worldwide demands an additional global production of 300 million metric tons (FAOSTAT 2015). Potato is one of the most valuable vegetable crops in Canada. Canada is one of the top 20 potato producing countries worldwide, with an annual production of over 4.7 million tonnes (STATCAN, 2016). Potato is a vital crop in terms of food security in the face of population growth and increasing hunger rates. For example, China, the world's biggest consumer of potatoes, expects to meet the demand to increase food production by 50% in the next 20 years (CIP, 2015). According to FAO statistics (2010), potato production in developing countries has increased by 94.6% over the last 15 years (Cromme et al., 2010). Out of the four major food crops such as rice, wheat, and maize, the potato has the best potential for high yield. The potato is a preferred crop in developing countries due to its nutritional value. The potato crop has wide adaptability to diverse environments and yield potential.

The potato crop is vulnerable to many biotic and abiotic stresses. Major diseases of potatoes include plant viruses (Hameed et al., 2014; Domfeh et al., 2014) nematodes (Mimee et al., 2015), bacteria (Salmond 1992) and fungi, including Chromista (Meng et al., 2009), causing substantial yield losses every year. The late blight of potato caused by *Phytophthora infestans* (an oomycete) is a severe production constraint in all potato producing countries. More than weekly applications of fungicides are needed to manage this disease. When the disease is severe, systemic fungicides are needed, for which pathogen often develops resistance (Ma and Michailides 2005). Despite this, usually under adverse weather conditions, the disease can destroy most of the crop in a week. The global loss due to this disease alone is estimated to be seven billion Can$ (Haverkort et al., 2009).

Breeding for resistance is considered to be a cost-effective and environmentally safe approach to manage late blight (Yoshida et al., 2013). Plant resistance against biotic stress is either qualitative or quantitative. Qualitative resistance, generally due to receptor *R* genes, is well characterized, and has been used in potato breeding (Watanabe 2015), but is often defeated by the pathogen producing a new effector. In contrast, quantitative resistance is durable, but breeding is challenging due to

polygenic inheritance. Further, the molecular and biochemical mechanisms underlying quantitative resistance are poorly understood (Kou and Wang 2010).

Resistance in plants against pathogen infection is defined as a spectrum of reduced susceptibility, ranging from susceptibility to a hypersensitive response or no symptoms (Eckardt 2002). Resistance in plants is primarily due to antimicrobial biochemicals, called phytoalexins, and/or the structures formed from them, which reinforce the cell walls to contain the pathogen to the initial infection area. Both of these lines of defence may be constitutive or induced following pathogen invasion (Kushalappa et al., 2016a). Pathogen perception leads to the induction of resistance-related (RR) proteins and RR metabolites. Both constitutive and induced RR metabolites are biosynthesized by metabolic pathway genes. Upon perception of the pathogen produced elicitors/effectors (EL/E), by the host membrane-localized elicitor/effector recognition receptors ($R_{ELRR/ERR}$), the downstream genes are triggered to mount a defence. However, the complete hierarchies of resistance (*R*) genes triggered by $R_{ELRR/ERR}$ genes, such as, phytohormones (R_{PHR}), MAP kinases (R_{MAPK}), transcription factors (R_{TFs}) that regulate the downstream genes that biosynthesize these resistance related metabolites (R_{RRM}) and RR proteins (R_{RRP}), to directly suppress or contain the pathogen, are not completely elucidated (Kushalappa et al., 2016a). The total resistance effect in a plant is due to several of these hierarchies of genes, each producing a set of RR metabolites in specific metabolic pathways. Several resistance genes have been identified in potato against late blight, and in wheat and barley against fusarium head blight caused by *Fusarium graminearum*, and their resistance functions have been proved (Kushalappa et al., 2016a). These R genes, or their orthologs, can be used to improve plant resistance against biotic stress in potato.

The commercial cultivars also have all the functional resistance gene hierarchies, except that a few are polymorphic, disabling them to biosynthesize the candidate RR metabolites. The polymorphism is due to hybridization, natural or proposed induced mutations, and gene transfer that lead to insertion-deletion (InDels) and SNPs (Cheng et al., 2015; Kushalappa et al., 2016b). Two genotypes with similar levels of resistance, may have different resistance mechanisms or genes. The functional R genes can be found in other commercial cultivars or a national germplasm collection. The non-functional (r) gene(s) or gene segments in a susceptible plant can be replaced

by a functional resistance gene(s) or gene segments to revive its ability to biosynthesize the RR metabolites, thus the resistance effects (Kushalappa et al., 2016b).

Advancements in the next-generation sequencing technologies integrated with tools like transcriptomics, proteomics, metabolomics, and association mapping studies have facilitated the identification of functional genes and their allelic variants. The advancements in functional genomics have helped in exploring the mechanisms involved in resistance. Several genes for biotic stress resistance have been identified, and their resistance functions have been proved by comparing gene expression and metabolic profiles of resistant and susceptible genotypes, in addition to disease severity and pathogen biomass (Yogendra et al., 2015a; Dhokane et al., 2016). Several technologies are being explored to carry out gene replacement of polymorphic genes with a functional copy of resistance genes. Two promising genome editing technologies are the transcription activator-like effector nucleases (TALEN) and clustered regularly interspersed short palindromic repeats CRISPR-Cas9 systems (Rinaldo and Ayliffe 2015). The genome editing, through gene knock-out or mutation and knock-in, replacement, or targeting, of crop plants, is evolving as a new era of opportunities, especially in molecular breeding. However, advancements in precise gene insertion and replacement are yet to come (Kushalappa et al., 2016b).

The CRISPR-Cas9 system is a promising tool and considered to be the most straightforward platform, which can be used to delete, insert, or replace gene sequences (Xie and Yang, 2013; Zheng et al., 2014). In this system, 20 nucleotide guide RNAs (gRNA) specific to the target site are designed to enable tailor-made genome modifications. The gRNA and Cas9 vectors are developed and introduced into the crop plants. The gRNA binds to the targeted genomic region, and the Cas9 enzyme will then cleave double-strand DNA breaks. Double-stand DNA repair results in precise modifications in the targeted region of the gene of interest via non-homologous end-joining (NHEJ). On the other hand, the double-strand break is repaired by homology-directed repair (HDR) mechanisms (Gratz et al., 2015; Zheng et al., 2014). There are several ways to deliver genome-editing reagents into plant cells. The programmable nucleases, along with gRNA, can be delivered into plant cells by using transfecting plasmids that encode them. These plasmids can be delivered into plant cells by using *Agrobacterium*-mediated transformation (Zhang et al., 2016a) and particle bombardment (Svitashev et al., 2016). Editing plant genomes without introducing

foreign DNA into cells may alleviate regulatory concerns related to genetically modified plants (Woo et al., 2015). The delivery of purified CRISPR-Cas9 ribonucleoproteins (RNPs) directly to the protoplasts of grape, apple, Arabidopsis, tobacco, lettuce, and rice plants has been reported (Woo et al., 2015; Malnoy et al., 2016). This approach may not be feasible for all plants and crops where the regeneration of plants from protoplasts is difficult (Burris et al., 2016). non-integrating plasmids could be transfected into plant cells to deliver programmable nucleases. However, transfected plasmids are degraded in cells by endogenous nucleases, and the resulting small DNA fragments are sometimes (very rarely) inserted at both on-target and off-target sites in host cells (Cho et al., 2014). Delivery of genome editing components through plasmids to plant cells is very efficient. Interestingly, genome editing reagents can be expressed transiently, which retains its function for a short time (Liang et al., 2017). Plants can be later regenerated from callus cells, which transiently expressing genome editing reagents (CRISPR-Cas9 nucleases and sgRNA).

Double-stranded breaks (DBSs) induced by nuclease are frequently repaired via NHEJ pathway which is error-prone and create unwanted insertion and deletion (indels) that often leads to loss of gene function (Belmaaza and Chartrand 1994; Puchta et al., 1996; Puchta et al., 1999; Steinert et al., 2016). On the other hand, DBSs repair via HDR repair leads to the gain of function gene replacement and targeting in the presence of a homologous DNA repair template. HDR mediated gene targeting is very much required for genetic studies to elucidate gene functions. HR mediated gene targeting, or gene replacement has been reported in plants and crops such as tobacco, rice, wheat, tobacco, and potato (Liang et al., 2017; Endo et al., 2016).

A homologous donor DNA template should be delivered to achieve HDR (Li et al., 2018; Dong et al., 2020). The efficiency of donor DNA template delivery and the abundance has an impact on HDR mediated repair. Gemini virus-based vector system or DNA replicons have been used to deliver CRISPR components along with the donor template. The use of geminivirus-based system significantly increased the gene targeting in various plants and crops such as tobacco, tomato, potato, wheat and rice (Baltes et al., 2014; Čermák et al., 2015; Butler et al., 2016; Gil-Humanes et al., 2017; Wang et al., 2017). Successful vector DNA-free genome editing in various plants is possible with this approach, as there is no integration of DNA or geminivirus into the host plant genome (Gil-Humanes et al., 2017).

1.1 GENERAL HYPOTHESIS

Resistance is primarily due to the accumulation of different specific pathway-related antimicrobial metabolites and their deposition to reinforce the cell walls to contain the pathogen to initial infection. Hierarchies of genes are involved in the production of each specific pathway related to metabolites. One or more such hierarchies of genes are required for the accumulation of enough amounts of metabolites in a genotype to achieve high levels of resistance under commercial production conditions. Potato genotypes with varying degrees of resistance against late blight also vary in their metabolome and transcriptome profiles due to allelic variations among hierarchies of resistance genes. Susceptible commercial potato varieties such as Russet Burbank may have missing set(s) of RR metabolites because of mutations in the required functional resistance genes. CRISPR-Cas9 mediated replacement of mutated alleles in elite cultivar (Russet Burbank) may lead to higher accumulation of RR metabolites, thus imparting increased resistance against late blight in potato.

1.2 OBJECTIVES

1. To identify R gene(s) that are functional in a late blight resistant genotype but are polymorphic (non-functional) in Russet Burbank cultivar, based on the metabolic-transcriptomics approach.
2. To sequence candidate R gene(s) in both resistant genotype(s) and susceptible cultivar to further confirm polymorphism.
3. To functionally validate the resistance effects of the polymorphic candidate gene(s) in Russet Burbank cultivar, based on genome editing, using the CRISPR/Cas9 system.

CHAPTER II: REVIEW OF LITERATURE

2.1 The fourth most important food crop: Potato

Potato (*Solanum tuberosum* L.) is the fourth staple food crop of the world, following maize, rice and wheat (FAO 2015), contributing to food and nutritional security. Potato is consumed by a billion people daily, grown in more than 178 countries. Potato cultivation is increasing in land area in the developing world as a cash crop for millions of farmers (Lutaladio et al., 2009) . The annual production of potato is over 300 million metric tons globally. North America production is 25 million metric tons; Canada produces 4.7 million metric tons of potato (STATCAN, 2016). Potato cultivation is distributed worldwide and is growing in a wide range of habitats. It ranges between 2000 and 4000 m altitude and from 47˚S to 65 ˚N latitudes. The total potato production area across the world is 19 million ha. North America cultivate potato in 6 million ha area.

The average yield of potato varies with growing conditions. In Africa, it ranges from about 13 t/ha and 50 t/ha in Western Europe and the USA. Patatin is one of the nutritionally balanced plant proteins known, a tuber glycoprotein, is a major storage protein of potato, which makes potatoes a second staple food crop after soybean for the amount of protein/ha (Liedl et al., 1987). The potato tubers are a rich dietary source of fibre, carbohydrates, high-quality proteins, vitamins and minerals, and antioxidants (Burlingame et al., 2009). Raw potato flesh and skin contents would generally be 2.4 g for dietary fibre, 19.7 g for vitamin C, 15.7 g for carbohydrates, 1.7 g for protein content, per 100 g of white raw potato (Singh and Kaur 2016); (as per USDA National Nutrient Database - https://ndb.nal.usda.gov/ndb/). Potato tubers are also known for the availability of minerals, with iron and zinc 0.52 mg and 0.29 mg, respectively, per 100 g. Potatoes are not just a starchy food, but it is known as a vegetable is a good source of antioxidant compounds, including polyphenols, carotenoids, and vitamins (Singh et al., 2009).

2.2 Potato late blight caused by *Phytophthora infestans*

Late blight, caused by the Oomycete *Phytophthora infestans* Mont. de Bary, is historically one of the most destructive of all potato (*S. tuberosum* L.) diseases, which originated in the central highlands of Mexico (Niederhauser et al., 1954). *P. infestans* can infect many Solanaceae members (Nowicki et al., 2012). Late blight is one of the most destructive diseases of potato worldwide, causing significant economic losses annually (Foolad et al., 2008; Nowicki et al., 2012). *P. infestans* remain a considerable challenge because of the abundant effector proteins produced to account for genome plasticity and pathogen population diversity (Fry 2008). The pathogen spreads through the coenocytic mycelium, which later forms elongated sporangiophores. The pathogen is a hemibiotroph, where it acts as a biotroph at the beginning of the infection and then becomes a necrotroph, producing several enzymes. The optimum temperature for sporangia germination is between 20-25°C, and from sporangia, the zoospores formed at a temperature between 10-15°C. The sporangia germinate in colder conditions to set free three to eight zoospores (asexual) (Melhus 1915). It has been shown that *P. infestans* sporangial dispersal occurred in flowing water or by wind, whereas motile biflagellate spores swim for an hour before encysting (Aylor et al., 2011). Leaf tissue infection is widespread in potato. Both sporangia and zoospores germinate, produce appressoria, and directly penetrate the leaf (Schumann and D'Arcy, 2000). Small necrotic spots appear on the host tissues two days post-infection.

Infection expands or coalesces at optimum temperature and 100% relative humidity, leading to blight development. The pathogen can also adopt sexual reproduction when both mating types (A and B) are available. Hormone exchange between bisexual thalli is a prevalent mechanism that triggers the formation of antheridia and oogonia to facilitate fertilization (Judelson 1997). Oospores of the pathogen can survive in the soil for years and withstand cold and desiccation by reaching dormancy. *P. infestans* oospores will also succumb to temperatures above 43°C (Fry and Goodwin 1997). Necrotic spots develop into blighted areas with whitish hyphae growing around them on the lower side of the leaf. The development of blight continues until infected leaves die. The pathogen will spread until the whole shoot is rotten, under favourable conditions. The pathogen would cause the most severe outbreak due to the spreading of zoospores in a humid

condition. Tuber infection mainly occurs through the washing down of spores from infected leaves (Agrios, 1988). The pathogen advances mostly in the peripheral tissues until the skin and flesh of tubers become soft and brown. Secondary infection aggravates the damage with severe tuber degradation even after tuber is harvested (De Lacy Costello et al., 2001). The infected tubers that are stored can be a potential source of inoculum for subsequent crops (Fry et al., 2013).

2.3 Late blight of potato and disease management

Late blight disease management involves two major strategies: avoiding sources of inoculum and avoiding favorable conditions. Usage of disease-free potato seed followed by the elimination of cull piles and volunteer plants can be effective ways to manage the onset of this disease. Crop rotation with non-solanaceous crops is advisable only when volunteer plants are destroyed. Careful field selection with good drainage and avoidance of humid zones may diminish disease-promoting conditions. Irrigation management to prevent moisture on foliage or flooded soil is very crucial. Unnecessary vine growth and wet foliage can be reduced by controlled nitrogen fertilization. Other major management strategies to lessen tuber infections are variation in cultivation and hilling that could avoid excessive moisture in the soil (Miller et al., 2006). Apart from the above-mentioned cultural measure, it is required to have chemical control. Seed treatment with fungicide (fluazinam, mancozeb) before planting and application of protectants before row closure by foliage is recommended where the disease is endemic. Late blight disease management often involves weekly applications of fungicides throughout the growing season. Systemic agents such as dimethomorph along with regular protective sprays may be used in rotation to delay or prevent fungicide resistance (Haverkort et al., 2009). The potato producers are very cautious due to the emergence of mefenoxam/metalaxyl resistance in the late 1980s, which hampered chemical control of late blight (Deahl et al., 1993).

Conventional late blight disease management strategies are quite expensive, and the multiple fungicide application would cause a total loss of 500 USD/ha/season (Cooke et al., 2011), while fungicide application is also connected with higher health risks (Axelstad et al., 2011). Potato producers and researches are motivated to pursue novel disease control strategies, including the use of resistant cultivars to avoid fungicide applications. Great efforts have been made to produce varieties with qualitative resistance, based on conventional or transgenic techniques (Nærstad et

al., 2007). The large potato germplasm is exploited to achieve qualitative and quantitative late blight resistance.

2.4 Qualitative and quantitative late blight resistance

The complete late blight resistance is a breeder's hope which requires intense study of genes implicated in resistance. Qualitative resistance of *P. infestans* is very well characterized and it mostly follows the classical gene-to-gene model (Flor 1971). Qualitative resistance is mainly due to receptor genes (surveillance genes). The receptor R genes trigger the downstream PR-proteins, Super oxides, toxic metabolites which eventually leads to the hypersensitive responses (HR) (Vleeshouwers et al., 2000). Several such receptor *R* genes have been isolated from different potato genotype (Malinovsky et al., 2014). These receptor *R* genes code for the NBS-LRR type cell membrane receptor proteins with intracellular signalling ability. The pathogen effectors targeted by NBS-LRR plant cell membrane proteins possess an RXLR translocation motif and a C-terminal domain (Nazarian-Firouzabadi et al., 2019; Haas et al., 2009). Effector proteins coded by *AVR* genes found in gene-sparse, repeat-rich zones of the *P. infestans* genome, which are the main reasons for rapid pathogen evolution. Pathogens generally overcome the single receptor *R* gene associated resistance (Raffaele et al., 2010), making qualitative resistance not durable.

Stacking of these receptor *R* genes have been introgressed into susceptible genotypes to improve resistance, but the pathogen overcomes such stacked resistance (Ballvora et al., 2002). The receptor *R* genes, however, are only surveillance genes that trigger downstream *R* genes to produce callose, a sugar polymer, around hyphae to form papillae (Malinovsky et al., 2014). The pathogen produces several elicitors (pathogen cell wall contents, enzymes, toxins, etc.) and these elicitors are recognized by host cell membrane receptors (Elicitor Recognition Receptors – ELRRs). Further, these elicitors trigger or induce various plant immune responses. Generally called as elicitor triggered immunity (ELTI). Elicitors are not race specific, rather very general to that pathogen. The hemi-biotrophic pathogen, P. *infestans,* generally suppresses the elicitor and produce very race-specific effector proteins. Plant immune response due to effector protein recognized by cell surface receptors (effector recognition receptors –ERRs) is called effector-triggered immunity (ETI) (Jones and Dangl 2006).

Following the pathogen invasion, both types of immune responses (ELTI/ETI) can trigger standard signalling components, such as reactive oxygen species (ROS), and mitogen-activated protein kinase (MAPK) cascades, which leads to transcriptional reprogramming and production of antimicrobial metabolites (Tsuda and Katagiri 2010).

On the other hand, the quantitative resistance is due to several hierarchies of genes that regulate the biosynthesis of RR metabolites. However, more elaborate studies are required on the regulation of biosynthetic pathway genes and the accumulation of RR metabolites. All hierarchies of genes, such as MAP kinases (MAPK) and transcription factors (TFs), which are triggered by ELRR/ERR implicated in the regulation of resistance-related (RR) metabolites, still to be elucidated (Kushalappa et al., 2016a). Biotic stress has been defined as the degree of susceptibility, and the mechanisms of resistance appear to be familiar to different host-pathogen systems (Kushalappa et al., 2016a). The resistance is due to the hierarchies of genes that produce metabolic pathway specific sets of metabolites. The resistance in plants is due to the cumulative effects of several downstream genes that biosynthesize RR metabolites and/or RR proteins, which are antimicrobial and reinforce cell wall (Kushalappa and Gunnaiah 2013). Several specific metabolic pathway genes have been reported to be involved in the biosynthesis of metabolites that reinforce the secondary cell walls, such as phenylpropanoid, terpenoid, fatty acids, and alkaloids (Pushpa et al., 2013; Yogendra et al., 2015b; Kumar et al., 2016a; Dhokane et al., 2016). The commercial susceptible cultivars also have these functional genes, except for a few that are polymorphic; replacing these can enable them to produce these metabolites, thus enhancing late blight resistance (Kushalappa et al., 2016b).

2.5 Transcriptomics

Transcriptome can be defined as the total of all the messenger RNA molecules expressed by the genes in a cell or a tissue of an organism. The study of the transcriptome is known as transcriptomics (Lai et al., 2012). Gene expression and respective messenger RNA abundance are very active in a cell or tissue. The transcriptome changes immediately with a response to cellular level variation. DNA arrays can be used most prominently to concurrently measure messenger RNA abundance for a given set of genes (Lockhart et al., 2000).

RNA sequencing (RNA-seq) is a high throughput tool developed to study the transcriptome of an organism (Wang et al., 2009). Various DNA microarray technologies have been used extensively to quantify the transcriptome, such as serial analysis of gene expression (SAGE), cap analysis of gene expression (CAGE), and massively parallel signature sequencing (MPSS). However, RNA-seq is a sequence-based approach has numerous advantageous over hybridization-based methods, which include: no reliance upon existing genomic sequences, novel gene/transcript identification, the broad dynamic range of detection, exact quantification, single nucleotide polymorphisms (SNPs) identification, and the discovery of transcript boundaries (splice variants), has become the technology of choice to study the transcriptome (Wang et al., 2009).

Whole-genome and transcriptome sequencing of potato (Xu et al., 2011; Hirsch et al., 2014) facilitated several genomic studies including RNA-seq, to study gene functions, differential gene expression related to varieties of stress responses, cellular mechanisms, and development (Gong et al., 2015; Cho et al., 2016; Gálvez et al., 2016; Goyer et al., 2015). Several genomics approaches have been used to identify the resistance mechanisms in the compatible and non-compatible interaction of potato *P. infestans* pathosystems (Gao and Bradeen, 2016; Massa et al., 2011; Burra et al., 2014). Transcriptomics studies and RNA-seq analysis in potato revealed the involvement of shikimate, flavonoid, lignin, and terpenoid biosynthetic pathways in resisting late blight disease (Yogendra and Kushalappa 2016).

2.5.1 Transcriptome sequence analysis and computational tools

Analysis of RNA-seq data with robust, efficient, and statistically principled algorithms is very crucial to obtain a piece of valuable information since transcriptome sequencing produces enormous sequence data (Trapnell et al., 2012). The complete annotation and quantification of all genes and their isoforms across samples is possible due to various potential computation tools (Garber et al., 2011). The analysis of RNA-seq data fall into two major types based on the availability of a reference genome for mapping raw sequence reads as (i) reference genome-guided sequence analysis and (ii) *de novo* transcriptome analysis. The reference guided sequence analysis is generally divided into three heuristic steps as (i) mapping reads to the reference genome, (ii) transcript assembly, and (iii) expression quantification. The mapping of reads to the reference genome is the dedicated step in genome-guided sequence analysis. Bowtie (Langmead et al., 2010)

and TopHat2 (Kim et al., 2013) are the most widely used read aligner computational tools to align the sequence reads to the reference genome efficiently. Owing to its limitations that Bowtie (Trapnell et al., 2012) cannot align reads that span introns, TopHat overcomes this limitation by aligning reads to the genome that contains significant gaps and detects transcript splice sites (Trapnell et al., 2012). The reference genome sequence of potato can be obtained for analyzing RNAseq data and assembling reads (Potato Genome Sequencing Consortium, 2011). Sequence files and other information can be obtained for two potato species: the heterozygous diploid *S. tuberosum* Group Tuberosum cultivar, RH89-039-16 (RH), and the doubled monoploid *S. tuberosum* Group Phureja clone DM1-3 (DM) (Xu et al., 2011). Recent updates on sequences were made available on the Spud DB genome browser (Hirsch et al., 2014). Cufflinks are one of the most widely used computational tools for transcript assembly, which assembles individual transcripts from the sequence reads mapped to the reference genome. Cufflinks can be used to assemble reads into different transcripts to identify splice variants. The individually assembled transcripts are then subjected to expression quantification. The number of reads aligned to the reference genome is proportional to transcript abundance. Cuffdiff software, which is a member of the cufflinks package, is the most widely used to determine transcript abundance (Garber et al., 2011). Several packages like DEGSeq, EdgeR, differential expression analysis of count data (DESeq), and Myrna are also commonly used for identifying differential gene/transcript expressions (Oshlack et al., 2010).

2.5.2 Identification of genomic variants from RNA-seq data

Genome variants or genomic variants can be classified based on their consequences and mutation types and based on their impact on protein function and structure (Pagani and Baralle 2004). Genomic variants (GVs) could be single-nucleotide substitutions, or small insertions or deletions (Pagani and Baralle 2004). The SNPs are DNA sequence variations appearing when a single nucleotide, such as – A, T, C, or G – in the genome differs between lines or individual (Altshuler et al., 2000). INDEL polymorphisms (InDels) can be defined as small insertions and deletions of nucleotides in the genome of an organism (Mills et al., 2006). Genomic variants may fall within the coding sequence of genes, the non-coding sequence of genes, or in the intergenic regions between genes (Pagani and Baralle 2004). GVs that change the sequence or expression of

genes are known to be associated with diseases. Variation in the DNA sequences of an organism can affect gene expression and are often associated with disease development, variation in pathogen response and development (Shastry 2009).

GVs, including SNPs and InDels, can change the amino acid sequences (nonsynonymous) or can be silent by not changing the amino acid sequence (synonymous) or simply occur in the noncoding regions (Shastry 2009). Synonymous mutations or variations do not affect the protein sequence. On the other hand, nonsynonymous variation or substitution can be missense and nonsense mutations (Al-Haggar et al., 2012; Cordovado et al., 2012). Missense mutations are known to code for different amino acids and often affect the protein function (Al-Haggar et al., 2012). A nonsense mutation is a type of point mutation in a DNA sequence that results in a premature stop codon and can completely truncate the protein (Cordovado et al., 2012). GVs that are not in protein-coding regions may still have consequences for gene expression, gene splicing, transcription factor binding, or the sequence of non-coding RNA (Li et al., 2014). Hence, identification of different types of GVs in genes and appropriate analysis of effects may assist in the understanding of their impact on gene function.

There are several analytical approaches and methods available for novel SNPs and GVs discovery. The methods available are: hybridization analysis, restriction fragment length polymorphism (RFLP), denaturating HPLC and gel electrophoresis, electrochemical analysis, single base extension, DNA sequencing (Altshuler et al., 2000). The other methods are capillary electrophoresis (Drabovich and Krylov 2006), mass spectrometry (Griffin and Smith 2000), single-strand conformation polymorphism (SSCP) (Tahira et al., 2009), etc. GVs and SNPs can also be assayed by high throughput Next-generation sequencing technologies (NGS) by sequencing the whole genome. High throughput NGS-based sequencing methods allow the identification of multiple SNPs in a specific genome when compared to other SNP genotyping methods (Jiang and Wong 2009). NGS-based SNP genotyping methods are currently available for all model species, but they strictly rely on the excellent reference genome (Pina-Martins et al., 2016). Ways to identify GVs and SNPs are currently well developed for model species but purely rely on the availability of a (functional) reference genome, and therefore cannot be applied to non-model species (Nielsen et al., 2011). However, NGS-based SNP genotyping and SNP calling are mostly

tailored for the whole genome (re-) sequencing experiments. But, in many cases, transcriptome sequencing (RNA-seq) can be used as a cheaper alternative to identify SNPs located in transcribed regions (Lopez-Maestre et al., 2016).

SNP detection in plants helps in understanding the relationship between genotype and phenotype and can be effectively applied in various plant breeding programs (Clevenger et al., 2015). There are numerous experimental approaches available to carry out SNP calling in plants. All the developed innovative methods use different alignment programs, variant calling software, and programs. Burrows-Wheeler Transform (BWT) (e.g., BWA and Bowtie) alignment programs that can be used along with several software tools for SNP calling in plants (Li and Durbin, 2009; Langmead et al., 2009). HISAT2 is another improved alignment program, which is an extension of BWT for graphs algorithm (Kim et al., 2015). The latest version of HISAT program has few advantages over the other programs for SNP or variant calling include: allows indels of any length about minimum alignment score (previously, the maximum length of indels was 3 bp), improved template length estimation of RNA-seq reads by taking introns into account when reads contain SNPs, the SNP information is provided as an optional field in the SAM output, no alignment penalty for mismatches, insertions, and deletions if they correspond to known SNPs, etc. (Kim et al., 2015). Output files from HISAT2 alignments can be usually stored in the sequence alignment/map (SAM) format (Li et al., 2009), which can be converted into binary (BAM) files. SAM and BAM files are processed for SNP and indel detection by various software programs such as SAMtools, Freebayes, and GATK (Li and Durbin, 2009; Rosenfeld et al., 2010; Schrider et al., 2011).

2.6 Metabolomics

Metabolomics is considered as the most potent tools for crop improvement due to its broad applications. Metabolomics is a study of all the metabolites of an individual organism under the set of genetic or environmental conditions. All the metabolites of an organism, the “metabolome” can be represented as an ultimate phenotype of a cell, which is subjected to change based on modulation of gene expression, protein function, and environmental signals. Accumulation of metabolites in plants greatly varies in response to various stresses in plants. Metabolic profiling has been extensively used to study stress biology in plants. Metabolic profiling is the identification

and quantification of metabolites in an individual organism at defined conditions, or subjected to variable environments (Shulaev 2006). Metabolites are much closer to phenotype; fingerprinting metabolic changes can be useful building a link between the genotype and phenotype when used combination with other 'omics' techniques such as genomics, transcriptomics, and proteomics (Fiehn, 2002; Hall et al., 2002; Hall, 2006). Metabolomics is a potential tool to dissect genotype-phenotype association and to investigate the functions of genes for various crop traits (Okazaki and Saito 2012). Metabolic profiling of an individual plant, tissue, organs, etc. allows comprehensive phenotyping of different crop traits, including disease resistance (Schauer and Fernie, 2006; Bollina et al., 2011; Kumaraswamy et al., 2011; Pushpa et al., 2013; Yogendra et al., 2014). However, the metabolomics technology has not well advanced to identify and quantify the metabolome but regardless of this significant advance have been made on stress resistance in plants. Integration of metabolo-transcriptomics approach in potato revealed that the specific group of metabolites belonging to hydroxycinnamic acid amides (HCAAs) were crucial in secondary cell wall reinforcement to contain the *P.infestans* at the site infection (Yogendra and Kushalappa 2016). Most importantly, high fold accumulation of HCAAs in potato resistant genotypes were reported in different studies (Yogendra et al., 2014b, 2015b).

2.6.1 Metabolite extraction and analytical platforms

It is imperative to have a very comprehensive and efficient protocol to extract all metabolites in plant tissue and also analytical platforms to detect all metabolites (Shulaev 2006; Kushalappa and Gunnaiah 2013). However, no single solvent can be used to extract all metabolites produced by plant system, which may lead to loosing important metabolites (Maloney 2004) Hence, different solvents or combinations of solvents are used for the extraction (polar, semi-polar, non-polar). Extraction protocols with varying concentrations of methanol and chloroform have been used. Very few efficient protocols have been established to extract a wide range of secondary metabolites from various plants and tissues (Shulaev, 2006; De Vos et al., 2007; Wen et al., 2014). However, extraction protocol using aqueous methanol was established to detect maximum metabolites (Bollina et al., 2010).

Besides, no single analytical platform can detect all the complex metabolites present in a sample. Nuclear magnetic resonance (NMR) is the best as it gives structural information. However, the

sensitivity of this is very low and thus cannot detect most of the RR metabolites that are in low concentration in plants and thus it limits the application of NMR in this field (Fernie and Schauer 2009). Gas chromatography-mass spectrometry (GC-MS) has been effectively used, but this detects only volatile metabolites (Lisec et al., 2006). Most of the RR metabolites are non-volatile; even the derivatization also is not very comprehensive. Liquid chromatography high resolution-mass spectrometry (LC-HRMS) is the most commonly used platform among all the above. Here the metabolite is separated based on the concentration of aqueous organic solvent (methanol) and following a softer ionization (electron spray and atmospheric), the negatively and positively charged ions are detected. The resolution of MS is very important as it also can separate metabolites. Low-resolution ones are linear ion traps, medium ones are a time of flight, and high-resolution MS are the orbitraps. In the latter, it is often used at >600 resolution, meaning rarely two metabolites are in one peak.

2.6.2 Computational frameworks: software and databases

Comprehensive and semi-comprehensive metabolome analysis leads to large datasets. Softwares such as MetAlighn (Lommen and Kools 2012), XCMS (Colin A. Smith et al., 2006), XCMS^2 (Benton et al., 2008), MetAlign, MZmine (Katajamaa et al., 2006; Pluskal et al., 2010), Markerlynx have been used for mass peak extraction and alignment raw output data across the samples (Vos et al., 2007).

Metabolite identification is another limitation, as not all the metabolites in plants have been identified. However, lately, there is an exponential increase in the number of reference libraries related to plant metabolites (Tohge and Fernie, 2010; Kushalappa and Gunnaiah, 2013). Several databases with commercial and publicly available libraries such as KEGG, METLIN, KNApSAcK, Lipid Maps, PlantCyc, MASS BANK, ReSpecT can be used for identification/annotation following analysis. The use of LC, with soft ionization techniques such as Electron spray ionization (ESI) or atmospheric pressure chemical ionization (APCI), has enabled us to identify and quantify semi-polar compounds. The advent of different ionization techniques, which results in a positive or negative mode ionization, helped to identify critical secondary metabolite groups such as phenolic acids, phenylpropanoids, flavonoids, alkaloids,

glycosylates, saponins and others (Vos et al., 2007). LC-HRMS approach has successfully been used in identifying resistance-related metabolites (Bollina et al., 2010).

2.7 Genome editing in Plants

Genome editing is a novel technique for precise and predictable modifications in the genomic DNA of cellular organisms. The genome-editing approach can fulfill the demands of basic plant research and genetic improvement of crops. These techniques provide a new capability for target mutagenesis and precise editing of plant genes, and control gene expression to study gene function and genetic improvement of crops. Induction of double-strand breaks (DSBs) in a specific chromosomal site with the use of artificial sequence-specific nucleases (SSNs) has the potential to modify genomes rapidly in a precise and predictable manner. Double strand breaks repaired by plants' inherent repair mechanisms such as error-prone non-homologous end joining (NHEJ) pathway and homology-directed repair (HDR). NHEJ is known to produce nucleotide insertions, deletions, and substitutions. On the other hand, HDR also can repair the DSBs if homologous donor templates are present at the time of DSB formation (Symington and Gautier 2011a). Sequence-specific nucleases such as zinc-finger nucleases (ZFNs) (Lloyd et al., 2005; Zhang et al., 2010; Zhang and Voytas, 2011; Petolino, 2015) and transcription activator-like effector nucleases (TALENs) (Boch et al., 2009; Moscou and Bogdanove, 2009; Christian et al., 2010; Zhang et al., 2013) have been successfully adopted in plant genome editing. The CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated protein9 nuclease (Cas9) is considered to be the more straightforward and most effective targeted genome editing tools of all above-mentioned tools and the same has been used extensively in crop improvement (Gao et al., 2015). Several reviews were described as the applications and advances of CRISPR-Cas9 systems in plants (Yin et al., 2017; Schindele et al., 2020). The CRISPR/Cas9 has been proved as a very well-established genome editing tool in human cells (Cong et al., 2013). The technology is then successfully adopted for plant cells, and the application of this technology was successfully tested in plant cells (Shan et al., 2014). CRISPR-Cas9 system was adopted from prokaryotic type II adaptive immune system for genome editing, which composed of 20 nucleotide guide RNAs (gRNA) and Cas9 protein (RNA guided DNA endonucleases) to negotiate PAM sequence (Protospacer adjacent motif) to successfully bind and cleave the target DNA sequence (Shah et al., 2013).

2.7.1 Gene delivery to plant cell and expression of genome-editing reagents

CRISPR and other genome-editing reagents can be delivered to plant systems in four major methods, such as i) *Agrobacterium*-mediated transformation, ii) gene gun, or a biolistic particle delivery system; iii) polyethylene mediated protoplast transfection; iv) plasmid-based method. Agrobacterium transformation is the most efficient, but the conventional method leaves T-DNA in the recipient plant, which is not acceptable for some regulatory authorities. Biolistic particle bombardment is good, but the efficiency is very low, and for replacement, the donor copies are considered not enough (Baltes et al., 2014). PEG method is excellent, but for many plant species, regenerations of plants from protoplast are very difficult or impossible (Burris et al., 2016). Recent studies have reported DNA-free plant genome editing by delivering pre-assembled CRISPR/Cas9 ribonucleoproteins to protoplasts of several crops, including lettuce, rice, Arabidopsis, grapevine, apple and wheat (Woo et al., 2015; Zhang et al., 2016). However, protoplast transformation and regeneration can not be widely adopted to all plants and crops as it is a challenging route to obtain transformed progenies. Further, isolation of protoplast and plant regeneration is labor-intensive, time-consuming is not applicable for high throughput applications (Arencibia et al., 1998).

Rrecently plasmids have been used to clone the genes which do not integrate the foreign DNA. Efficient plasmid-based genome editing methods are available where CRISPR-cas9 and sgRNAs can be expressed transiently. CRISPR-Cas9 can be delivered as DNA or RNA (Nekrasov et al., 2013; Shan et al., 2014; Zhang et al., 2016). The plasmid-dependent delivery system for CRISPR-Cas9 and sgRNA transient expression (functional only for a short period) has been reported (Lawrenson et al., 2015; Zhang et al., 2016). These transient expression plasmids harboring CRISPR-Cas9 and gRNAs can be delivered either by agrobacterium or biolistic method. Plants can be later regenerated from callus cells, which transiently expressing genome editing reagents-CRISPR-Cas9 nucleases. In the recent advancement, it was also reported the possibilities of delivering gene-editing reagents through de novo meristem induction in a few plant species, including potato (Maher et al., 2020). However, there's a great hindrance to the development of the right strategies and delivery system in potato to achieve efficient homozygous mutants targeting all alleles (multiallelic) (Kusano et al., 2018). Because most of the cultivating potato genotypes are heterozygous and are tetraploid, these challenges are addressed efficiently over time

in various ways. Over time, successful multiallelic (tetra-allelic) gene-editing was reported in many studies with the widely available SNP physical mapping and other progress in choosing the specific target sequences (Veillet et al., 2019a; Enciso-Rodriguez et al., 2019; Sevestre et al., 2020). These studies were proved the efficiency of the CRISPR-Cas9 tool for carrying out functional genomic studies in potato by knocking-out genes in all alleles.

2.7.2 Gene replacement and Gemini virus-based gene targeting in plants

Targeted mutagenesis or CRISPR-based gene knock-out via NHEJ has been demonstrated widely in several transformable plant species. NHEJ is still the most preferred repaired DNA repair pathway in plants. On the other hand, gene targeting, or precise modification is challenging due to ineffective methods for delivering donor homologous repair template for achieving efficient HDR based repair. However, independently replicating viral-based vectors have been successfully used for providing genome-editing reagents in model plants and crops such as tobacco, potato, tomato, rice and wheat (Baltes et al., 2014; Čermák et al., 2015; Butler et al., 2016; Gil-Humanes et al., 2017; Wang et al., 2017). Geminivruses (family *Geminiviridae*) are single-stranded DNA (ssDNA) viruses that can infect a wide range of plant species like wheat, maize, cotton, tomato, legumes, etc. (Rey et al., 2012). Geminiviruses can replicate inside plant cells through homologous recombination-dependent replication and rolling circle replication, which makes them an excellent system for gene targeting or gene replacement (Baltes et al., 2014). To date, two ssDNA plant viruses have been used for genome engineering or gene targeting in plants: Bean yellow dwarf virus (BeYDV) in potato and tomato; wheat dwarf virus (WDV) in cereals such as rice and wheat (Zaidi and Mansoor 2017). Importantly, there is no integration of DNA or geminivirus into the host plant genome that takes place with related vectors (Gil-Humanes et al., 2017).

CONNECTING STATEMENT FOR CHAPTER III

As reviewed in chapter II, the late blight disease caused by the oomycete *Phytophthora infestans* is still a severe disease and causes yield loss and reduces tuber quality. Several management strategies have been used to manage the disease, including fungicide application, resistance genotype use, other cultural and biological practices. All the available methods have their drawbacks and difficulties in applying. Developing the late blight resistant genotype is the best way to manage the late blight disease. The quantitative resistance is reduced susceptibility to the pathogen and mainly due to the resistance-related metabolites (Kushalappa et al., 2016a). Importantly, these metabolites can render a durable host resistance by blocking the pathogen at the site of infection due to the reinforcement of the cell wall. Resistance metabolites and conjugates were identified in the late blight resistance potato cultivars, and they were mainly derivatives of the phenylpropanoids, flavonoids, and fatty acids (Pushpa et al., 2013; Yogendra and Kushalappa 2016). All the cultivated potato cultivars are highly heterozygous, and the many traits, including disease resistance, are affected because of the deleterious mutations (Xu et al., 2011; Zhang et al. 2019). Here it was proposed that the deleterious mutation among the biosynthetic genes can alter the resistance metabolite production in the late blight susceptible cultivar Russet Burbank. In the following chapter, the study was conducted to identify SNPs and InDels among biosynthetic genes that produce resistance metabolites responsible for cell wall reinforcement during the infection. In the study, the deleterious genes were identified among three biosynthetic pathways, including phenylpropanoid, flavonoid, and fatty acid. All the SNPs/InDels were categorized as deleterious based on their impact on protein structure and function. Besides, the few of the critical deleterious mutations were confirmed by Sanger sequencing the respective gene fragments of the Russet Burbank genotype. Biosynthetic genes that were polymorphic between susceptible (Russet Burbank and resistance (F06037) were selected. Subsequently, the qRT-PCR experiment was to confirm their expression in pathogen and mock (water control) Russet Burbank samples.

CHAPTER III

Identification and functional characterization of late blight resistance polymorphic genes in Russet Burbank potato cultivar

Running title: Disease resistance genes polymorphic in Russet Burbank potato

Niranjan Hegde[A], *Dadakhalandar Doddamani*[B], *Ajjamada C. Kushalappa*[A,]

[A]Department of Plant Science, McGill University, Ste.-Anne-de-Bellevue, QC, Canada

[B]The Roslin Institute, The University of Edinburgh, United Kingdom

3.1 Abstract

In plants, the biosynthesis of the phenylpropanoid, flavonoid, and fatty acid pathway monomers, polymers and conjugated metabolites play a vital role in disease resistance. These are generally deposited to reinforce the cell walls to contain the pathogen to the site of infection. Identification of sequence variants in genes that biosynthesize these resistance metabolites can explain the mechanisms of disease resistance. The resistant and susceptible genotypes, inoculated with *Phytophthora infestans*, were RNA sequenced to identify the single nucleotide polymorphisms (SNPs) and insertion/deletion (InDel) variations. The SNPs/InDels were annotated and classified into different categories based on their effect on gene functions. In selected 25 biosynthetic genes overlapping 39 transcripts, a total of 52 SNPs/InDels were identified in the protein-coding (CDS) regions. These were categorized as deleterious based on the prediction of their effects on protein structure and function. The SNPs/InDels data obtained in this study can be used in genome editing to enhance late blight resistance in Russet Burbank and other potato cultivars, if mutated, to enhance disease resistance.

Keywords:

RNA-seq, phenylpropanoids, flavonoids, fatty acids, metabolic pathways, cell-wall reinforcement, SNPs/InDels, deleterious mutations, late blight of potato, disease resistance.

3.2 Introduction

The late blight caused by an oomycete pathogen *Phytophthora infestans* is one of the most devastating diseases of potato worldwide. The disease is managed by weekly fungicide applications, but often the pathogen causes complete crop loss during rainy and humid conditions; also poor farmers cannot afford to spray and thus the loss in their fields is always higher (Haverkort et al., 2009; Forbes et al., 2019; Arora et al., 2014). In addition, the use of systemic fungicides leads to the buildup of resistant populations of the pathogen, making the fungicide applications useless (Ma and Michailides 2005). Genetic improvement of the crop is an excellent option but the sexual incompatibility in potato makes it very challenging to breed cultivars (Mendoza et al., 1996).

Resistance in plants, innate immune response, is considered to be reduced susceptibility (Wiesner-Hanks & Nelson, 2016; Kushalappa et al., 2016a). The pathogens are classified into biotrophs and necrotrophs based on their food habits. The necrotrophs following inoculation of host plants produce elicitors that trigger Ca^{2+} and reactive oxygen species (ROS), which induce downstream genes to produce programmed cell death (PCD) or molecular pattern triggered immunity (PTI). The biotrophs generally suppress their elicitor production, enter the cell wall and produce more specialized effectors, which trigger downstream genes to produce hypersensitive response type of PCD (HR-PCD) or effector-triggered immunity (ETI). The elicitors and effectors are recognized by the host membrane-localized pattern recognition receptor (PRR) proteins or genes (receptor *R* genes). Several receptor *R* genes have been pyramided to enhance late blight resistance (Vossen et al., 2016). However, the pathogen overcomes such resistance as the receptor *R* genes are only surveillance genes and they must be associated with the downstream *R* genes that can confer resistance. These receptor *R* genes are also involved in triggering pathogen-related signalling (Zaidi and Mansoor 2017). The signal transduction genes such as MAP kinases (MAPKs) and signalling genes that produce phytohormones further trigger the regulatory genes such as

transcription factors (TFs), which eventually regulate the biosynthesis of several resistance-related (RR) metabolites and proteins that suppress the pathogen progress, thus the resistance (Kushalappa et al., 2016b).

Plants produce more than 200 000 metabolites, including monomers, polymers and conjugated metabolites which are antimicrobial and/or are deposited inside the cell walls to contain the pathogen to the initial infection area (Yogendra et al., 2015b; Kumar et al., 2016a; Dhokane et al., 2016). These RR metabolites are biosynthesized in different metabolic pathways, such as phenylpropanoid, flavonoid, fatty acid, terpenoid, and alkaloid. The polymers and conjugated forms, such as hydroxycinnamic acid amides (HCAAs), monolignols, lignans, wax, and suberins reinforce the cell walls preventing further progress of the pathogen (Wang et al., 2013). These complex compounds and polymers cannot be easily broken down by the enzymes secreted by pathogens. Variations in the levels of resistance in plants, thus, depend on the type, number and amount of these RR metabolites produced by that plant. These RR metabolites are biosynthesized in metabolic pathway networks involving biosynthetic *R* genes and as well as the regulatory, signalling and receptor *R* genes, following pathogen invasion (Kushalappa et al., 2016a). SNPs/InDels in these hierarchies of genes can disable the biosynthesis of a given metabolite(s), thus making the resistant plant susceptible to a given pathogen. We hypothesize that the SNPs/InDels of RR metabolite biosynthetic *R* genes are rendering the most commonly cultivated Russet Burbank potato cultivar susceptible to late blight, and editing of these genes to make them functional would enable this cultivar to biosynthesize RR metabolites, which can suppress the pathogen and reduce disease severity.

Variant calling or SNPs/InDels identification from RNA-seq, instead of whole-genome sequencing, could decrease the experimental costs and are also useful in discovering SNPs/InDels in transcribed regions of genes, for studying direct functional/structural impact on the protein function (Lopez-Maestre et al., 2016). In recent years, attempts have been made in other plant species, where the RNA-seq was used for high-throughput variation calling analysis (Rogier et al., 2018; Zhao et al., 2019). NGS based variant calling in potato has been used to discover SNPs/InDels (Uitdewilligen et al., 2013; Mosquera et al., 2016).

This is the first report on the identification of SNPs/InDels associated with the secondary cell wall reinforcement by RR metabolite biosynthetic *R* genes against late blight in potato. In this study, multiple annotators have been used to comprehensively categorize variations, mainly to target the transcribed and protein-coding regions, which can impact on protein structure. We have identified RR metabolites that are low in abundance in late blight susceptible Russet Burbank cultivar relative to two late blight resistant genotypes. RNA seq was analyzed in both Russet Burbank and two resistant genotypes, following inoculation with *P. infestans* (Yogendra and Kushalappa 2016) The SNPs/InDels were identified and were associated with the transcripts coding the metabolite biosynthetic enzymes and as well the abundances of RR metabolites biosynthesized. The SNPs/InDels were categorized into different types based on their impact on protein structure and functions. The plausible role of these SNPs/InDels and their impact on RR metabolite biosynthesis and resistance to late blight is discussed. A few of these *R* genes and their respective mutations were validated in susceptible and resistant cultivars by Sanger sequencing and gene expression studies. These genetic variations identified in gene transcripts including *StCCoAOMT* can be corrected to restore resistance alleles in Russet Burbank potato cultivar to biosynthesize these RR metabolites, thus enhancing resistance to late blight.

3.3 Materials and methods

3.3.1 SNPs/InDels analysis in RNA-seq reads

The RNA-seq data on late blight resistant genotypes (F06025 and F06037) and susceptible commercial cultivar Russet Burbank (RB) was obtained from our previous study (Yogendra and Kushalappa 2016). Briefly, the leaves were inoculated with the mock solution and *P. infestans* sporangial suspension, and the leaf discs containing the inoculated leaf area were sampled at 2 dpi. The Illumina short reads from resistant and susceptible cultivars were aligned to the *Solanum tuberosum* L. Group Phureja genome as reference genome (The Potato Genome Sequencing Consortium 2011) using HISAT2 default parameters, with a maximum of two mismatches (Kim et al., 2015). The output file was saved in SAM format. SAMtools was used for variant calling or to predict SNPs and InDels within reliable and mapped reads (Li et al., 2009; Li, 2011), followed by SnpEfftool (Cingolani et al., 2012) and SIFT 4G (Vaser et al., 2015) were used separately, to annotate effect of SNPs and InDels into different categories. The effect of variants on gene

functionality was later categorized as described earlier (Cingolani et al., 2012; Vaser et al., 2015). The number of heterozygous and homozygous mutations were calculated using SnpEff output files. In this study, three annotators were separately used to annotate SNPs/InDels variations into different categories. The SnpEff output annotation file was used to sort out genotype-specific variations and to compare SNPs/InDels variations among all three genotypes. SNPs/InDels specific to Russet Burbank were classified and separately interpreted (Fig. 3.1, 3.2a & 3.2b).

3.3.2 Identification of SNPs/InDels in metabolite biosynthetic genes and assessment of deleterious mutations

The *S. tuberosum* phenylpropanoid, flavonoid and wax biosynthesis pathways genes were downloaded from phytazome (https://phytozome.jgi.doe.gov/pz/portal.html) and KEGG databases. (https://www.genome.jp/kegg-bin/show_organism?org=sot). The genetic variations (SNPs/InDels) within the biosynthetic pathway genes were discovered in an annotated output variants (files) obtained from SnpEff and SIFT 4G tool. All the annotated SNPs/Indels annotated in metabolite biosynthetic genes were further confirmed and interpreted by using the Ensembl Variant Effect Predictor (VEP) online tool (http://plants.ensembl.org/Tools/VEP?db=core). Consequences of non-synonymous mutations and missense mutations within the protein-coding region were assessed utilizing SIFT and PROVEAN online tools. Mutations were categorized into HIGH, MODERATE, MODIFIER, and LOW. HIGH impact mutations mainly include premature stop codon, frameshift, and stop lost mutations types. The mutations were considered as deleterious because it is confident that they can affect protein structure and function. A missense mutation was categorized as a MODERATE type. Missense mutation can be a substitution of one base for another in an exon of a gene coding for the protein. Missense mutations were further named as synonymous and non-synonymous substitutions. Synonymous substitutions are silent mutations without changing the amino acid sequences. In contrast, non-synonymous substitutions alter the amino acid sequences of the protein.

3.3.3 Validations in SNPs/InDels based on PCR, sequencing, and qRT-PCR

Genes and respective SNP/InDels identified were validated in RB and F06037. The details for primers, chromosome locations, gene name, and PGSC (Potato Genome Sequencing Consortium) IDs are presented (Table 3.1). All the PCR amplified samples were Sanger sequenced to confirm the variations discovered in variant calling using RNA-seq data. Gene expression analysis (qRT-PCR) was performed on pathogen inoculated leaf samples collected at 2 dpi, as described in RNA sequencing, using IQ SYBR Green Supermix (BioRad) in a CFX384 Real-Time System (BioRad). Affinity Script qRTPCR cDNA Synthesis Kit (Agilent Technologies, Stratagene Products Division, La Jolla, CA, USA) was used to synthesize cDNA and from each sample, 3mg purified RNA was collected. For qRT-PCR, PrimerQuest Tool (Integrated DNA Technologies) was used to design primer sets for transcripts with HIGH impact SNPs/InDels (Table 3.2 & 3.3), wherein *St EF1α* and *stb-tubulin* were used as a reference gene (Nicot et al., 2005).). The relative gene expression level was calculated using the 2–DDCt method (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008).

3.3.4 Characterization of the effects of deleterious SNPs/InDels on amino acid composition and *In Silico* analysis to validate protein conformational changes at the structural level

A total of 11 non-synonymous deleterious SNP variations identified in RB, were selected and analyzed further to characterize their effect on protein structure. Peptide and CDS sequences were predicted, using protein-based gene prediction tools FGENESH+ (Solovyev 2007), for mutant and native allelic variants. The amino acid sequence obtained for transcript template (mutant and native alleles) were then used to predict 3D protein structure using I-TASSER (Yang et al., 2015). 3D protein structures were visualized and carried out superposition (Match -> Align, a tool in the Structure Comparison) by Chimera1.13.1 (https://www.cgl.ucsf.edu/chimera/) (Meng et al., 2006). The predicted 3D protein structure was also analysed using missense3D (Ittisoponpisan et al., 2019), to examine the effect of amino acid substitution on protein structural changes.

3.3.5 Semi-targeted metabolic profiling

The resistance-related (RR) metabolites previously identified in resistant genotypes, F06037 and F06025 were used here (Yogendra and Kushalappa 2016). Briefly, the leaves inoculated with the mock solution and *P. infestans* and samples were collected at 3 dpi. The metabolites were extracted using acqueous methanol and analyzed using liquid chromatography and high-resolution mass spectrometry (LC-HRMS, LC-Orbitrap MS, Thermo scientific). The peaks were annotated using mzMine software. Monoisotopic masses with significantly higher abundances in resistant than in susceptible genotypes were considered as resistance-related (RR) metabolites. These were further grouped into RR constitutive (RRC=RM/SM>1.0) and RR induced (RRI=[(RP/RM>1.0)/(SP/SM>1.0)], R is resistant, S is susceptible genotype; P is a pathogen and M is mock solution inoculated.

3.4 Results

3.4.1 Transcriptome sequences of late blight resistant and susceptible potato genotypes

Transcriptome sequence data were obtained from late blight resistance (F06025 and F06037) and susceptible (Russet Burbank) potato genotypes inoculated with *P. infestans.* After assembly, 59.58, 58.76, 68.09, million paired-end reads (100 bp) were obtained in F06025, F06037, and RB, respectively (Table 3.4). SAMtool was used to analyze RNA-seq data for variant calling (SNPs/InDel identification). More than 850000 SNP variations and 36000 InDel variations were discovered in all three potato genotypes (Table 3.5). The workflow diagram of the entire analysis of the study is shown (Fig. 3.3). The SNPs/InDel distribution plot was very dense and it explained a uniform distribution of SNPs/InDels across 12 chromosomes of all three potato genotypes. Accordingly, the distribution of SNPs/InDels identified among phenylpropanoid, flavonoid, and fatty acid biosynthetic pathway genes of RB were plotted across 12 chromosomes (Fig. 3.4a, 3.4b). More than 63% of SNPs identified were heterozygous with 66.18% in F06025, 63.55 in F06037 and 68.81% in RB (Table 3.6). However, the percentage of heterozygous InDels in all the three genotypes were more than 90%, with 91.35% in F06025, 91.32 in F06037, and 91.71% in RB (Table 3.6).

SnpEff data was used to classify, the SNPs/InDels based on their overall impact on gene function into LOW, MODERATE, MODIFIER, HIGH impacts in all three genotypes. The HIGH impact SNPs included the stop-gain, stop-lost, frameshift, etc. known to have direct impacts on gene functions due to protein truncation, degradation of the transcript and different translations from the origin. The HIGH impact mutations identified within biosynthetic pathway genes were mainly focused. Few of the genes, with HIGH impact InDels identified in RB, were grouped based on biological functions and their association with disease resistance. The LOW impact SNPs/InDels mainly comprised of the synonymous and splice region variants within the protein-coding regions. Synonymous or silent substitutions have little or no impact on proteins because these within genes are known to unalter the amino acid sequence. The MODIFIER SNPs/InDels fall within the intergenic, intronic and intragenic regions. Particularly, intragenic SNPs/InDels within 3′-UTR and 5′-UTR regions can alter the binding sites for transcription factors (TFs) and miRNAs. As a result, the variations can alter the gene expressions. But the present study mainly emphasized the variations within protein-coding regions (CDS).

In addition to SnpEff annotator, the SIFT 4G annotator was also used to analyze SNPs, as the protein-coding region variants can also have a direct impact on protein functionality. The SIFT 4G annotator was used to predict the non-synonymous amino acid substitutions (MODERATE SNPs/InDels) and their effect (deleterious or tolerated) on protein functions based on SIFT score. Using the annotator, 32 904 non-synonymous variations (MODERATE SNPs/InDels) were categorised (protein-coding regions) into deleterious variations. SIFT 4G did not show any predictions for HIGH SNPs/InDels within the protein-coding regions of genes because the HIGH impact variations like stop-gained, stop-loss and frameshift mutations change the protein sequence. So, the HIGH impact mutations were not considered for further assessment to predict the protein structure and comparison of mutant and native forms.

3.4.2 Identification of deleterious mutations in genes mainly involved in the biosynthesis of metabolites that reinforce secondary cell walls

Phenylpropanoid pathway genes: A total of 3303 SNP variations overlapping within 445 transcripts of 272 genes, related to phenylpropanoid biosynthetic pathway were identified in RB

(Fig. 3.5a). Likewise, the variations in 163 InDels covered 34 transcripts of 27 genes. The HIGH impact SNPs/InDels with stop-gained (<1%), stop-lost (<1%), frameshift (<1%) variations were considered as main candidates, and were further discussed to explore their role in the differential accumulation of pathogen-induced metabolites in resistance and susceptible potato genotypes. The deleterious mutations were spotted on transcript variants of those genes which are involved in defense-related suberisation and lignification. Transcript variants of three major genes including caffeoyl-CoA O-methyltransferase (*StCCoAOMT)*, and Class III peroxidases (*StPrx*) were identified with deleterious mutations (Table 3.7). Notably, these transcript variants are known to have a crucial role in the reinforcement of the plant cell wall in response to pathogen challenge, through increased production of cell wall-bound ferulic acid polymers and related compounds. The gene and protein names of all biosynthetic genes discussed in this study are presented (Table 3.8). O-methyltransferase family 2 protein-coding genes were also identified with the stop-gain mutation. The transcript variant of this gene is well known for its role in melatonin and quercetin biosynthesis (Table 3.7). Further, genes and transcripts with, missense or non-synonymous SNPs (42%) and InDels (>1%) variations were picked, to assess their impact on protein structure and gene functions. Very fine screening led to the identification of deleterious missense or non-synonymous mutations in a total of 18 transcript variants of nine different genes (Table 3.9). Only nonsynonymous substitution with >0.5 SIFT score and >-2.5 PROVEAN score were chosen for further analysis (Table 3.9). Transcript variants with deleterious missense mutations were observed in genes: 4-coumarate--CoA ligase encoding genes (*St4CL1* and *St4CL2*), cinnamoyl-CoA reductase (*StCCR*), cinnamyl alcohol dehydrogenase encoding genes (*StCAD1, StCAD6, StCAD9*). Only a few genes and respective transcripts of Orcinol O-methyltransferase and Omega-hydroxypalmitate O-feruloyl transferase (*StASFT1*) were identified with InDel variations.

Flavonoid pathway genes: Since flavonoids and their accumulation play a significant role in plant disease resistance, flavonoids biosynthesis pathway genes were explored. A total of 1375 variations were identified in 265 transcripts of 159 genes involved in flavonoid biosynthesis. The percentage of SNPs/InDel variation distribution in genes involved in the flavonoid biosynthetic pathway varied from 55.2% to 36.24% (Fig. 3.5b). A total of 24 genes were identified with 151 InDels. The HIGH impact SNPs/InDels were identified within NAD(P)H oxidoreductase, isoflavone reductase transcript variants (Table 3.7). Transcript variants of this gene are known to

have a role in phytoalexin (Pterocarpan) biosynthesis. Protein isoforms produced from a few other transcript variants of the same gene take part in lignan (8-8'-linked lignans, pinoresinol, and lariciresinol) biosynthesis. Interestingly, a few transcripts of chalcone synthase (*StCHS2*) and flavonol synthase/flavanone 3-hydroxylase (*StFLS1*) genes were also discovered with stop-loss and frameshift mutations (Table 3.7). The transcript of feruloyl CoA ortho-hydroxylase 1 (*StF6'H1*), a gene that biosynthesizes scopoletin was found mutated in RB. Deleterious missense and non-synonymous SNPs/InDels variations within other genes (Table 3.9) were selected to carry out protein structure prediction.

Fatty acid biosynthetic genes: Phenylpropanoid thioesters (Feruloyl Co-A and Caffeyl-Co-A) can conjugate with unsaturated fatty acids including long-chain fatty acids and fatty alcohols (C18–C30), as well as hydroxy fatty acids and dicarboxylic acids (C14–C20) (Kosma et al., 2012; Domergue et al., 2010). Feruloylation is a key process in producing alkyl ferulates and alkyl caffeates during wound healing and defense-related suberin. Mutations or SNPs/InDels variations in these could affect the biosynthesis of unsaturated fatty acids (wax biosynthesis) which later can impede defense-related suberisation. A total of 483 SNPs within 60 transcripts of 37 genes and 28 InDels within 15 transcripts of 9 genes were identified in the fatty acid pathway. Deleterious mutations were discovered within the transcript variants of Acyl-CoA--sterol O-acyltransferase 1(*StASAT1*), feruloyl-Coenzyme A transferase like (*StASFT*) and sterol desaturase (*StCER1*) (Table 3.7 & 3.9).

3.4.3 Russet Burbank mutated gene validation in late blight resistant F06037 genotype

To gain confidence in SNPs/InDels discovered here, a few of these mutated genes in RB were validated for their disease resistance functions. A total of 10 stop-gain HIGH impact SNPs/InDels corresponding to 10 genes in the phenylpropanoid pathway were further confirmed to be polymorphic, based on polymerase chain reaction (PCR) and Sanger sequencing, in RB and F06037. Transcripts of all 10 genes with HIGH impact mutations were highly expressed in pathogen inoculated leaves of resistance genotype (F06037) as compared to susceptible RB (Table 3.7; Fig. 3.6).

3.4.4 Analysis of the effect of non-synonymous mutations on amino acid compositions of protein structures

Polymorphism in alleles of metabolite biosynthetic genes can lead to varying levels of metabolite production and disease resistance. To assess the role of mutated genes in cell wall reinforcement and as well in disease resistance, the 3D protein structures were predicted, for native and mutant alleles based on amino acid sequences, especially in the phenylpropanoid pathway (Table 3.9). For example, the SIFT and PROVEAN scores were validated by predicting protein structures for a transcript of three native wild type cinnamoyl-CoA reductase (CCR^{A79}, $CCR^{S205, W226}$, $CCR^{S54, W75}$) and respective variant alleles (CCR^{A79T}, $CCR^{S205F, W226R}$, $CCR^{S54F, W75R,}$) where the amino acids are indicated in superscripts, and their positions are given in superscript and the mutant alleles with amino acid substitutions in italics. In addition, the 3D protein structures were predicted for variant and native alleles of PAL1 and 4CL1 (PGSC0003DMT400037490 (a); PGSC0003DMT400037485 (b)); ($PAL1^{A3, C535, I489}$; $PAL1^{A3S, C535Y, I489V}$) and $4CL1^{Y54, V189, G383}$; $4CL1^{Y54H, V189M, G383D}$). The Protein superposition revealed the protein structural variation and conformational changes, especially in the secondary structure of proteins between mutated and native forms (Fig. 3.7a-3.7f). The structural divergence between native and variant forms for all the alleles was clear. A few of the results obtained after superposition, using matchmaker and match-align tools in UCSF Chimera 1.13.1 including Root-mean-square deviation of atomic positions (RMSD) values, were confirmed by missense 3D (Table 3.10). The missense 3D results for CCR^{W226R} (PGSC0003DMT400001401), $PAL1^{C535Y}$ (PGSC0003DMT400049886) and $4CL1^{Y54H}$ (PGSC0003DMT400037485; PGSC0003DMT400037490) transcript variants (variant alleles) were further analyzed. Hence, speculated the role non-synonymous mutations within biosynthetic genes and their association with high susceptibility of Russet Burbank to late blight disease. However, the protein structural variations among the late blight resistance and susceptible potato genotypes and their precise role in late blight resistance require a detailed study in the future.

3.5 Discussion

Resistance in plants to pathogen stress is mainly due to RR proteins and metabolites (Kushalappa et al., 2016b). Plants produce thousands of RR metabolites, but not all are required to achieve the level of resistance needed for commercial cultivation. Two cultivars with equal levels of resistance

may have different resistance mechanisms or different sets of metabolites, each involving a hierarchy of genes for their biosynthesis. A cultivar has less resistance because some of these RR metabolites biosynthetic or their regulatory genes are mutated (allelic variation), thus unable to biosynthesize them. Here RNA seq was used to identify the SNPs/InDels of these *R* genes and their application in enhancing the levels of late blight resistance in Russet Burbank cultivar is discussed.

Variation calling using RNA-seq and genome-wide next-generation sequencing (NGS) methods have been previously reported to associate SNPs to potato late blight resistance and have been used as diagnostic markers (Mosquera et al., 2016; Enciso-Rodriguez et al., 2018). The RNA-seq analysis also has been used to identify differentially expressed genes (DEGs) and association with late blight disease resistance in various potato genotypes (Gao et al., 2013; Yogendra & Kushalappa, 2016; Yang et al., 2018). The present study centred on RNA-seq analysis to identify variations (SNPs/InDels) in pathogen-induced transcripts of late blight resistant and susceptible genotypes.

The approach to identify SNPs/InDels variations calling used here was similar to those used in previous reports in different crops (Vidal et al., 2012; Rogier et al., 2018; Zhao et al., 2019). Three different annotation tools were separately used, not only to categorize mutations into different types but also to predict the strong functional effects of every SNP/InDel variation. A high percentage of heterozygosity with respect to SNPs/InDels variations was observed in all three genotypes studied here. High heterozygosity in cultivated potato genotypes has been previously reported (Uitdewilligen et al., 2013). With a great number of studies available in different crop plants, variation calling has become a very useful tool for genetic mapping and SNP application in plant trait improvement (Clevenger et al., 2015). The SNPs/InDels variations have been successfully used in SNP calling tools and to do annotations to categorize variations into different types (Li et al., 2016; Ramakrishna et al., 2018). In the present study, a total of 54 deleterious mutations were identified in 39 transcripts of 25 genes belonging to phenylpropanoid, flavonoid, and fatty acid biosynthesis pathway (Table 3.7 and 3.9). Some of these polymorphic genes and alleles are involved in the biosynthesis of polymer and conjugated metabolites that reinforce the cell walls to contain the pathogen to the initial infection area (Matern et al., 2011; Miedes et al.,

2014; Kushalappa et al., 2016). Suberisation and lignification are the two major processes involved in the secondary cell wall thickening. Since resistance is mainly due to reinforcement of secondary cell walls by the RR metabolites that polymerize or conjugate with other metabolites, the SNPs/InDels in genes that biosynthesize them are discussed below. The metabolomics information previously published on these genotypes in our lab is used to substantiate the impact of genes with SNPs/InDels identified in RB (Yogendra and Kushalappa 2016).

3.5.1 Hydroxycinnamic acid amides (HCAAs)

HCAAs accumulation was more abundant in resistance genotypes (F06025, F06037) than in RB genotype, upon pathogen inoculation (Yogendra and Kushalappa 2016). Similar studies have also been carried out in different crops, where the importance of HCAAs synthesis in defense response have been implicated (Muroi et al., 2009). We reported elevated accumulation of N-feruloyltyramine N-feruloyloctopamine, feruloylputrescine, feruloylagmatine in resistant genotype F06037 as compared to RB (Yogendra and Kushalappa 2016). Higher expression of genes and deposition of HCAs conjugated with amines have also been identified in RB in several of our previous studies (Yogendra et al., 2015; Yogendra & Kushalappa, 2016). Interestingly, all HCAAs identified were mainly feruloyl-CoA conjugates. Moreover, HIGH impact SNPs/InDels identified in (*StCCoAOMT*) gene encoding, caffeoyl-CoA O-methyltransferase (methylates caffeoyl-CoA to feruloyl-CoA and 5-hydroxyferuloyl-CoA to sinapoyl-CoA) can be the main reasons for less HCCAs production in RB. Interestingly, protein network analysis (https://string-db.org/ and https://genemania.org/) have revealed the plausible close interactions of StCCoAOMT protein with transcript sequences of *StPAL1I, St4CL1, St4CL2,* (upstream to *StCCoAOMT* gene in phenylpropanoid pathway) genes. However, deleterious mutations in transcripts of these genes further strengthened the main cause for reduced HCCAs formation in RB, following pathogen invasion. Further experimental proof is required to get insights into enzyme-enzyme interactions and their implications in metabolic pathway regulation. Similar interactions among other candidate genes have been proved and additional mechanisms have been explained in various plant-pathogen interactions (Bassard et al., 2012; Biała & Jasiński, 2018). Mainly, 4-coumarate--CoA ligases are encoded by the multigene family and produce CoA thioesters, which is a branch point directing metabolites to flavonoid or monolignol pathways (Wang et al., 2016). However, our study identified deleterious mutations within only a few transcripts (of *St4CL1* and *St4CL2*), which are

convincing and in accordance with previous studies describing their exclusive role in cell wall reinforcement in response to pathogen invasion (Li et., 2015; Yang et al., 2018; Chen et al., 2019). The protein structural variations in these candidates also can greatly impact enzyme-substrate binding. In this study, only protein structures were predicted for few deleterious variations that fall within the protein coding regions (Fig 3.7a-3.7f).

3.5.2 Wall bound ferulic acid polymers, feruloylated polysaccharides, and wax related suberin synthesis

Two transcripts of caffeoyl-CoA O-methyltransferase (*StCCoAOMT*) (PGSC0003DMT400006134 and PGSC0003DMT400006135) were identified with stop-gain mutations in Russet Burbank. The orthologue of this gene in maize conferred resistance against multiple pathogens (Yang et al., 2017). However, the gene isoforms are known to have a very specific role in the synthesis of wall-bound ferulic acid polymers and feruloylated polysaccharide in potato (Grimmig & Matern, 1997; Nakane et al., 2003). Having the same role, along with *StCCoAOMT,* frameshift mutations in transcripts of two class III peroxidase and InDels identified in Feruloyl-Coenzyme A Transferase like (*StASFT*) gene transcripts may greatly affect the production of cell wall-bound ferulic polymers and feruloylated polysaccharides. The distinct roles of different types of class III peroxidases (Almagro et al., 2008) and acyltransferases encoded by *ASFT* genes (Rautengarten et al., 2012) in plant defense have been reported. In addition, SNPs/InDels variations were also discovered in transcripts of three acyl-CoA dependent acyltransferase (BAHD-type and HXXXD-type) in phenylpropanoid pathway (PGSC0003DMG400001075, PGSC0003DMG400009273) and fatty acid biosynthesis pathway (PGSC0003DMG402013278). The detailed functions, mainly ferulylation of fatty acids and suberin synthesis in different plant tissues, of BAHD and HXXXD type acyltransferases have been reported (Gou, Yu, & Liu, 2009; Molina et al., 2009). A few of the SNPs in these transcripts were validated by PCR and Sanger sequencing. In one of our previous studies, we have reported a higher accumulations of (15Z)-12-oxophyto-10,15-dienoicacid, (9Z,11E,15Z)-(13S)-hydroperoxyoctadeca-9,11,15-trienoate and PC(16:0/2:0) (Fold Change=3.78) metabolites in F06025 genotype and (9Z,12Z)-(8R)-hydroxy octadeca-9,12-dienoic acid, (9Z,11E,15Z)-(13S)-hydroperoxyoctadeca-9,11,15-trienoate (17.90) and PC(O-16:0/3:0) (2.96) metabolites in F06037 than in RB (Yogendra and Kushalappa 2016). Hence, we speculate that the mutations in the above-

mentioned peroxidases and acyltransferases encoding genes may affect the ferulylation of these fatty acids and aromatic suberization in RB genotype. Based on these results and previous reports the putative resistance mechanism is schematically presented (Fig. 3.8a & 3.8b) and the details on mutation type, location, transcript and gene IDs are presented.

3.5.3 Monolignols and lignan biosynthesis

Phenylpropanoid-derived lignin and monolignol biosynthetic genes are induced in response to many plant pathogens. Lignification is regarded as one of the common phenomenon in plants to contain the invading pathogen (Dixon & Paiva, 1995; Q. Zhao & Dixon, 2014). Generally, the phenylpropanoid pathway is divided into two steps: (i) phenylalanine to feruloyl CoA and (ii) feruloyl CoA to monolignol. (Fig. 3.8a and 8b) (Vanholme et al., 2013). Cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD) catalyze the final designated steps in monolignol biosynthesis (monomers of lignin polymers and lignans). Identification of protein structural variations (Fig. 3.7a-3.7f) due to the deleterious mutations within specific transcript variants belonging to CCR (PGSC0003DMG401018219, PGSC0003DMG400025373, PGSC0003DMG400000521) and CAD (PGSC0003DMG400021152, PGSC0003DMG400018446, PGSC0003DMG400016984) encoding genes explain their preferential expression during the biotic stress, leading to regulation of cell wall reinforcement. The CCR and CAD genes, which are multigene encoded as 4CL and upregulated in the expression of few gene isoforms during plant-pathogen interaction, have widely been reported (Lauvergeat et al., 2001). The *StCCR* and *StCAD* were highly upregulated in resistant genotypes than in the RB genotype (Yogendra et al., 2015b), even though we did not find a higher accumulation of phenolic monolignols in the resistance genotypes. Interestingly, non-synonymous mutations were identified (Table S4) in a few transcripts of a gene encoding PHENYLCOUMARAN BENZYLIC ETHER REDUCTASE 1, which belongs to the NAD(P)H oxidoreductase, isoflavone reductase and can synthesize 8-8'-linked lignans, pinoresinol, and lariciresinol. Altogether, these results suggested that a good number of CCR and CAD encoding transcript variants are involved in the biosynthesis of RR monolignols.

3.5.4 Flavonoids

Flavonoids also are very important phytoalexins and antimicrobial compounds (Yuan et al., 2008; Lu et al., 2017). A higher accumulation of scopoletin has been reported in resistant genotypes than in RB (Yogendra and Kushalappa 2016). However, *StF6'H* gene which codes for feruloyl CoA ortho-hydroxylase 1 is found mutated (frameshift mutation) in RB. Further, it was speculated that the accumulation of some of the very important RR flavanols such as kaempferol and quercetin seems to be hindered by the mutations in flavanol synthase/flavanone 3-hydroxylase encoding genes and their transcripts. Flavonol synthase (*StFLS*) or flavanone 3-hydroxylase (*StF3H*) is a key enzyme that catalyzes dihydro flavanols to flavanols, including quercetin, kaempferol, and myricetin. Our previous study showed the accumulation of kaempferol and quercetin related compounds (flavonol glycosides) in late blight resistant genotypes (F06037 and F06025), following pathogen inoculation (Pushpa et al., 2013; Yogendra et al., 2014). As the antimicrobial activity of flavonol glycosides have been widely implicated in disease resistance (Tagousop et al., 2018; Chen et al., 2019), our previous study also reveals an abundant accumulation of cyanidin 3-O-rutinoside kaempferol 3-(200-p-coumaryl-rhamnoside)-7-rhamnoside and flavonol 3-O-b-D-glucosyl-(1- >2)-b-D-glucosyl-(1->2)-b-D-glucoside in the F06025 genotype, and 1-O-vanilloyl-b-D-glucose, cyanidin 3-O-(6-O-p-coumaroyl)glucoside and cyanidin 3-Orutinoside in F06037, as compared to RB genotype, following pathogen invasion (Yogendra and Kushalappa 2016). Biosynthesis and accumulation of flavanols sustain the plants from ROS damage, following pathogen and other stresses (Nakabayashi et al., 2014). The isoflavone reductases (IFR) identified here are also crucial enzymes involved in the synthesis of phytoalexins, such as prenylated pterocarpans (glyceollins) (Oliver et al., 2003; Cheng et al., 2015). These phytoalexins are known to accumulate in response to biotic and abiotic stresses (Wang et al., 2013). In this study, deleterious mutations in transcripts of these genes were identified in the RB genotype (Table 3.7 & 3.9).

3.5.5 Validating the effects of mutations on protein structure and disease resistance functions

A significant reduction in the expression of several metabolite biosynthetic genes was found in RB, including some with HIGH impact SNPs/InDels as compared to resistance genotype, following pathogen inoculation (Fig. 3.6a-3.6j). Some transcripts with stop-gain mutations and frameshift mutations were found to result in a truncated transcript or the production of a completely different translation from the origin. These types of mutations alter the stop codons leading to an abnormally short or long polypeptides and transcripts, which are often not functional (Griffiths et al., 1999; You et al., 2007). Most importantly, the aberrant transcripts (mRNA) accumulation are generally inhibited by nonsense-mediated decay (NMD) in eukaryotes including plants (Wu 2007; Shaul, 2015). There was a lower accumulation of transcripts bearing stop-gain and frameshift mutations in RB, even after the pathogen inoculation (Fig. 3.6a-3.6j). Further investigations are required to better understand the role of NMD-targeted transcripts in disease response (Rayson et al., 2012). Interestingly, the outcome of the expression results also gave a clue on the resistance functions of all the transcripts we have identified with deleterious mutations. Only non-synonymous missense mutations in transcripts were further studied to assess the impact on protein structure.

The present study identified not only the functional *R* genes but also several important SNPs/InDels variations, mainly in the phenylpropanoid-flavonoid and fatty acid pathway metabolite biosynthetic *R* genes in one commercially grown potato cultivar, Russet Burbank, and in two resistant genotypes. The SNPs/InDels in a commercial cultivar were identified, not by comparing to resistant genotypes, but rather based on the functionality of *R* genes, which were classified into different groups based on the impact of mutations on protein production. Based on this we have identified SNPs/InDels not only in late blight susceptible genotype but also in resistant genotypes. The importance of these mutations on late blight resistance can be validated in RB cultivar by editing these *R* genes and quantifying disease severity and pathogen biomass. Identification of RR metabolites in edited and non-edited plants can prove the resistance functions. Though a couple of late blight resistant genotypes are used here, the functional genes for editing

can also be obtained from several other cultivars and landraces in germplasm collections. Metabolic pathway regulation is very complex and a step by step identification and editing of these genes eventually can attain high levels of resistance required for the commercial production of Russet Burbank cultivar, with a minimum of other management options.

Table 3.1: Details of primers used in this study to validate putative SNPs by PCR and Sanger sequencing.

Gene Name/transcript name	Ref/Var	Location	PGSC ID	Primer sequence (Forward and Reverse)
HXXXD-type acyl-transferase protein	G/T	ch11_13690558	PGSC0003DMG400009273	5´-TGTTATCTTTCATTTTTGAGACGCT-3´
				5´-TGCCTCAACAAGCCATGGAA-3´
Aminotransferase	A/T	ch06_59410014	PGSC0003DMG400024037	5´-TGCTTCGCAAGTACAGGGAG-3´
				5´-TGCAAAAACAACTAGCTCTCCTC-3´
Orcinol-o-methyltranferase	G/C	ch06_15007066	PGSC0003DMG401025689	5´-GTAGTCACTATAACCAGCATCAAAA-3´
				5´-GATATTACATGATTGGAGCGACGG-3´
Flavonol synthase	A/G	ch02_30627265	PGSC0003DMG400033906	5´-TCTTGGACCGTCAAGGGATG-3´
				5´-TCATTAACCTGAAGGAGTATGGCT-3´
Chalcone synthase	T/G	ch09_58364068	PGSC0003DMG400029621	5´-ATTGGGCCTAAAGCCTGAGA-3´
				5´-TTTCGAAAGATCCTCGGCCC-3´
Wax synthase	T/G	ch11_572732	PGSC0003DMG402013278	5´-TGATGTTGGACCGAATCCCG-3´
				5´-GACATCTGGCTTACCCCTCA-3´
Caffeoyl-CoA O-methyltransferase	C/A	ch04_57240328	PGSC0003DMG400002387	5´-GCTAGAATCTTGCCATCATCAGG-3´
				5´-CTTCACAAGCATTGACGGACA-3´
Phenylalanine ammonia-lyase 2	G/T	ch09_5539942	PGSC0003DMG401021564	5´-GCAGAAGTTGAGGCAAGTCC-3´
				5´-GGAGCACCATTCCAGCTCTT-3´
Trans-caffeoyl-coa 3-o-methyltransferase	G/T	ch08_1766514	PGSC0003DMG401026272	5´-CTGCTACAACTGGTGCGTTG-3´
				5´-AGGCACCTTAACAAGGCAAG-3´
Cinnamyl alcohol dehydrogenase 7	A/C	ch11_1411969	PGSC0003DMG400016985	5´-TCTTGCAGCTCTTTCCCACA-3´
				5´-ACAGCGTTGGAACGTCTTCT-3´

Table 3.2: Transcript IDs and primer details for the qRT-PCR experiment, used in gene expression analysis of transcripts (with HIGH impact mutations) in pathogen inoculated F06037 and Russet Burbank leaf samples**.**

AssaySet	Type	Sequence	Start	Length	Tm	GC Percent
a PGSC0003DMT400029815	Forward Primer	5′-AACATCCCAGGACCAAACTC-3′	1019	20	61.918	50
a PGSC0003DMT400029815	Reverse Primer	5′-CACCTGACAGAGTGACCATATC-3′	1112	22	61.732	50
b PGSC0003DMT400006135	Forward Primer	5′-CTCTGCTTCTCACTGCTCTTAC-3′	831	22	62.108	50
b PGSC0003DMT400006135	Reverse Primer	5′-CTGGCAGGCCTAGTTCATAAG-3′	920	21	62.094	52.381
c PGSC0003DMT400039064	Forward Primer	5′-TTGTTGGCAAAGCCAGAAATAA-3′	464	20	61.196	50
c GSC0003DMT400039064	Reverse Primer	5′-GCAAGAGGAAGGAGATAGTGAAG-3′	567	22	60	36.4
d PGSC0003DMT400007964	Forward Primer	5′-GCAAGAGGAAGGAGATAGTGAAG-3′	464	23	62.024	47.826
d PGSC0003DMT400007964	Reverse Primer	5′-GAGTAGGGAGAGTGATCATGTAAAG-3′	567	25	61.876	44
e PGSC0003DMT400007990	Forward Primer	5′-CTGTTCGCTGTCTCTGATGAA-3′	385	21	62.001	47.619
e PGSC0003DMT400007990	Reverse Primer	5′-CCGTGTACTCTTCAATCGTCTC-3′	520	22	62.04	50
f PGSC0003DMT400076179	Forward Primer	5′-GGGTGGCCGCAGTTATTAT-3′	653	19	62.084	52.632
f PGSC0003DMT400076179	Reverse Primer	5′-CGCTATCAGGGACTAGAGTTTG-3′	751	22	61.717	50
g PGSC0003DMT400084273	Forward Primer	5′-GTTGGTATCCTTGTGGGAGAAG-3′	822	22	62.271	50
g PGSC0003DMT400084273	Reverse Primer	5′-GAACCGAGGCTTCTTGTATGT-3′	919	21	62.114	47.619
h PGSC0003DMT400007961	Forward Primer	5′-GCAAGAGGAAGGAGATAGTGAAG-3′	464	23	62.024	47.826
h PGSC0003DMT400007961	Reverse Primer	5′-GAGTAGGGAGAGTGATCATGTAAAG-3′	567	25	61.876	44
i PGSC0003DMT400074657	Forward Primer	5′-CGAGCCATCGTTTCTTCAATTT-3′	345	22	61.901	40.909
i PGSC0003DMT400074657	Reverse Primer	5′-CTCTCCATCCCAATGTCTTCTC-3′	473	22	61.821	50
j PGSC0003DMT400034544	Forward Primer	5′-CTTGGTGGCACTACTTCTTTCT-3′	273	22	62.351	45.455
j PGSC0003DMT400034544	Reverse Primer	5′-GAAAGTGGTAGAGGTGGGATTG-3′	379	22	62.271	50

Table 3.3: Primers used in gene expression (qRT-PCR) experiments.

GenBank name	Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
AB061263.1	*StEf1-α*	ATTGGAAACGGATATGCTCCA	TCCTTACCTGAACGCCTGTCA
Z33382.1	*Stβ-tubulin*	ATGTTCAGGCGCAAGGCTT	TCTGCAACCGGGTCATTCAT

Table 3.4: Raw data processing of RNA seq data in potato genotypes.

Filename	HI.2303. 008.Index_18. F06037		HI.2303. 008.Index_14. Russet Burbank		HI.2303. 008.Index_7. F06025	
	R1	R2	R1	R2	R1	R2
Total bases	5,876,675,800	5,876,675,800	6,809,785,400	6,809,785,400	5,958,721,700	5,958,721,700
Total reads	58,766,758	58,766,758	68,097,854	68,097,854	59,587,217	59,587,217
% bases >=Q20	97.485	92.94	97.548	93.427	97.597	93.34
% bases >=Q30	92.73	86.709	92.891	87.3	93.058	87.265
Average read length	100	100	100	100	100	100
Read length range	100 .. 100	100 .. 100	100 .. 100	100 .. 100	100 .. 100	100 .. 100
Quality range	35 .. 73	35 .. 73	35 .. 73	35 .. 73	35 .. 73	35 .. 73
Phred range	2 .. 40	2 .. 40	2 .. 40	2 .. 40	2 .. 40	2 .. 40
Offset	33	33	33	33	33	33
A	1,717,714,857	1,718,899,728	1,953,615,707	1,986,309,751	1,743,317,054	1,731,017,660
T	1,699,539,477	1,707,702,585	1,961,317,488	1,935,083,868	1,714,803,830	1,732,024,958
G	1,125,787,459	1,323,275,581	1,343,690,959	1,545,989,934	1,144,932,406	1,343,586,057
C	1,333,261,315	1,124,790,059	1,550,727,543	1,340,081,796	1,355,289,323	1,150,057,033
N	372,692	2,007,847	433,703	2,320,051	379,087	2,035,992
percent G-C content	41.844	41.657	42.504	42.381	41.959	41.849

Table 3.5: Total number of SNPs/InDels identified in potato genotypes based on the SnpEff annotation tool**.**

Total number of SNPs identified in each genotype			
Genotype	**Total**	**Homozygous**	**Heterozygous**
F06025	894687	336415	558272
F06037	858217	320546	537671
Russet Burbank	858217	323378	555661
Total number of InDels identified in each genotype			
Genotype	**Total**	**Homozygous**	**Heterozygous**
F06025	38446	3323	35123
F06037	36164	3136	33028
Russet Burbank	36867	3054	33813

Table 3.6: Types and positions of SNP and InDel variations in three potato genotypes. The number is based on SnpEff and SIFT 4G annotation tools.

Variant type	Number	Positions	Number
frameshift deletion	9291	cds	613173
frameshift insertion	9711	utr3	248711
noncoding	371105	utr5	141581
nonframshift deletion	192	others	581
nonframshiftinsertion	129		
nonsynonymous	270603		
synonymous	336481		
start-lost	503		
stop-gain	1949		
stop-loss	526		
substitution	3556		

Frameshift: Genetic mutation caused by InDels in a DNA sequence that is not divisible by three and change the reading frame
Stop-lost: Point mutation within the termination codon and lead to abnormal extension of the protein at the carboxyl terminal
Stop-gain: Point mutation within the protein coding region results in premature stop codon and resulted in short protein
Non-synonymous: Substitution of one base for another in an exon of a gene coding for a protein without changing any amino acid sequence.
Synonymous: Substitution of one base for another in an exon of a gene coding for a protein by changing amino acid sequence.
cds: Coding region of a gene
utr: untranslated region

Table 3.7: Genes with deleterious HIGH impact mutations detected in phenylpropanoid, flavonoid, and fatty acid-wax biosynthesis pathways. Genes were identified following *Phytophthora infestans* inoculation in susceptible genotype Russet Burbank. Chromosome number, location, allelic variations, transcripts IDs, variation types, and SIFT prediction scores are given in detail. The variable type is showed as SnpEff annotator**.**

Chrom	Pos	Transcript_id	Gene_name	Mutation_type	Variation
1	47297343	PGSC0003DMT400029815	Class III Peroxidase	Framshift variant	p.Ala13fs
4	57240328	PGSC0003DMT400006135	Caffeoyl-CoA O-methyltransferase	Stop_gained	p.Glu80*
4	57240328	PGSC0003DMT400006134	Caffeoyl-CoA O-methyltransferase	Stop_gained	p.Glu42*
11	11931662	PGSC0003DMT400039064	Peroxidase 5 cell wall	Framshift variant	p.Ser243fs
11	11931662	PGSC0003DMT400039067	Peroxidase 5 cell wall	Framshift variant	p.Ser243fs
2	30627265	PGSC0003DMT400007964	Flavonol synthase/flavanone 3-hydroxylase	Stop_gained	p.Arg9*
2	30594268	PGSC0003DMT400007990	Flavonol synthase/flavanone 3-hydroxylase	Stop_lost&splice_region_variant	p.Ter217Cysext*?
2	30627265	PGSC0003DMT400007964	Flavonol synthase/flavanone 3-hydroxylase	Stop_lost&splice_region_variant	p.Ter209Trpext*?
9	58364068	PGSC0003DMT400076179	Chalcone synthase 2	Stop_lost&splice_region_variant	p.Ter380Gluext*?
10	44291854	PGSC0003DMT400084273	NAD(P)H oxidoreductase, isoflavone reductase	Stop_lost&splice_region_variant	p.Ter246Glyext*?
2	30627607	PGSC0003DMT400007961	Flavonol synthase/flavanone 3-hydroxylase	Framshift variant	p.Cys299fs
11	10118479	PGSC0003DMT400074657	Feruloyl CoA ortho-hydroxylase 1	Framshift variant	p.Glu192fs
11	572732	PGSC0003DMT400034544	Wax synthase isoform 3	Stop_gained	p.Tyr268*

* stop codon. Ala- Alanine, Glu- Glutamic acid, Ser- Serine, Arg- Arginine, Gly- Glycine, Cys- Cysteine, Tyr- Tyrosine, fs- frameshift,

Table 3.8: Metabolite biosynthetic genes and protein names in phenylpropanoid, flavonoid, and fatty acid biosynthetic pathway.

Gene ID	Protein	Gene
PGSC0003DMG400011458	Class III Peroxidase	*StPrx*
PGSC0003DMG400015106	Peroxidase 5 cell wall	*StPrx*
PGSC0003DMG401018219	cinnamoyl-CoA reductase	*StCCR*
PGSC0003DMG400025373	cinnamoyl-CoA reductase	*StCCR*
PGSC0003DMG400021152	cinnamyl alcohol dehydrogenase 9	*StCAD9*
PGSC0003DMG400000521	Cinnamoyl-CoA reductase	*StCCR*
PGSC0003DMG400018446	cinnamyl alcohol dehydrogenase	*StCAD*
PGSC0003DMG400002387	caffeoyl-CoA O-methyltransferase	*StCCOAOMT*
PGSC0003DMG401025689	Orcinol O-methyltransferase	*StCCOAOMT*
PGSC0003DMG400028929	4-coumarate--CoA ligase 2	*St4CL-2*
PGSC0003DMG400001075	Omega-hydroxypalmitate O-feruloyl transferase	*StASFT*
PGSC0003DMG400014466	4-coumarate--CoA ligase 1	*St4CL-1*
PGSC0003DMG400015677	Sinapyl alcohol dehydrogenase-like 1	*StCAD1*
PGSC0003DMG400015678	cinnamyl alcohol dehydrogenase 6	*StCAD6*
PGSC0003DMG400016984	Cinnamyl alcohol dehydrogenase 7	*StCAD7*
PGSC0003DMG400019386	Phenylalanine ammonia-lyase	*StPAL*
PGSC0003DMG400003079	Flavonol synthase/flavanone 3-hydroxylase	*StFLS1*
PGSC0003DMG400011929	flavonol synthases like	*StFLS- like*
PGSC0003DMG400019110	Chalcone Synthase	*StCHS2*
PGSC0003DMG400020471	Isoflavone reductase homolog	N/A
PGSC0003DMG400028505	2-oxoglutarate (2OG) and Fe (II)-dependent oxygenase	N/A
PGSC0003DMG400026630	Malonyltransferase	*StPMAT2*
PGSC0003DMG400033906	NAD(P)H oxidoreductase, isoflavone reductase	N/A
PGSC0003DMG400029028	Feruloyl CoA ortho-hydroxylase 1	*StF6'H1*
PGSC0003DMG400013347	Acyl-CoA--sterol O-acyltransferase 1	*StASAT1*
PGSC0003DMG402013278	Feruloyl-Coenzyme A Transferase like	*StASFT*
PGSC0003DMG401023841	Sterol desaturase	*StCER1*

Table 3.9: Genes with deleterious non-synonymous missense mutation detected in a) phenylpropanoid pathway, b) Flavonoid pathway, c) fatty acid biosynthesis pathway. Genes were identified following *Phytophthora infestans* inoculation in susceptible genotype Russet Burbank. All the details including chromosome number, location, allelic variations, gene IDs, variation types, are given. Variants with the SIFT (Sorting Intolerant From Tolerant) score in the range of 0.0 to 0.05 were considered "deleterious" and a PROVEAN (Protein Variation Effect Analyzer) score less than or equal to −2.5 was considered "deleterious".

Transcript id	Gene name	Variation	Sift score	Provean score
Phenylpropanoid pathway				
PGSC0003DMT400046910	cinnamoyl-CoA reductase	V208M	0.034	-3.6
PGSC0003DMT400054502	cinnamyl alcohol dehydrogenase 9	S172C	0.013	-4.7
PGSC0003DMT400001401	Cinnamoyl-CoA reductase	S205F	0.025	-3.4
PGSC0003DMT400001400	Cinnamoyl-CoA reductase	S54F	0.045	-3.4
PGSC0003DMT400001400	Cinnamoyl-CoA reductase	W75R	0	-13.5
PGSC0003DMT400001401	Cinnamoyl-CoA reductase	W226R	0	-13.4
PGSC0003DMT400047475	cinnamyl alcohol dehydrogenase 9	D130N	0.037	-4.6
PGSC0003DMT400074444	4-coumarate--CoA ligase 2	S164C	0.011	-3.6
PGSC0003DMT400049886	Phenylalanine ammonia-lyase	C535Y	0.043	-6.2
PGSC0003DMT400043757	Cinnamyl alcohol dehydrogenase 7	E64K	0.001	-3.7
PGSC0003DMT400040519	Sinapyl alcohol dehydrogenase-like 1	R104C	0.006	-3.7
PGSC0003DMT400040518	Sinapyl alcohol dehydrogenase-like 1	R222C	0.014	-4.2
PGSC0003DMT400040518	Sinapyl alcohol dehydrogenase-like 1	L176F	0.034	-3.2
PGSC0003DMT400040519	Sinapyl alcohol dehydrogenase-like 1	L58F	0.044	-3.1
PGSC0003DMT400037490	4-coumarate--CoA ligase 1	Y54H	0.018	-3.8
PGSC0003DMT400037485	4-coumarate--CoA ligase 1	Y54H	0.019	-3.9
PGSC0003DMT400037486	4-coumarate--CoA ligase 1	G279D	0.006	-5.5

PGSC0003DMT400037485	4-coumarate--CoA ligase 1	G383D	0.012	-5.6
PGSC0003DMT400037490	4-coumarate--CoA ligase 1	G383D	0.013	-5.5
PGSC0003DMT400037489	4-coumarate--CoA ligase 1	G43D	0.039	-6.1
Flavonoid pathway				
PGSC0003DMT400007963	Flavonol synthase	F204I	0.018	-5.0
PGSC0003DMT400007961	Flavonol synthase	F204I	0.022	-4.9
PGSC0003DMT400007962	Flavonol synthase	F16I	0.044	-5.2
PGSC0003DMT400031148	flavonol synthases like	L216F	0.021	-3.5
PGSC0003DMT400052746	Isoflavone reductase homolog	R303C	0.003	-2.5
PGSC0003DMT400052746	Isoflavone reductase homolog	K74N	0.006	-4.3
PGSC0003DMT400084271	Isoflavone reductase	G12S	0.02	-5.0
PGSC0003DMT400084272	Isoflavone reductase	G12S	0.02	-4.8
PGSC0003DMT400084273	Isoflavone reductase	G12S	0.023	-5.0
PGSC0003DMT400084271	Isoflavone reductase	H67P	0.002	-7.5
PGSC0003DMT400084272	Isoflavone reductase	H67P	0.002	-7.8
PGSC0003DMT400084273	Isoflavone reductase	H67P	0.002	-8.1
PGSC0003DMT400084272	Isoflavone reductase	S256L	0.027	-2.5
PGSC0003DMT400084272	Isoflavone reductase	P293L	0.001	-9.4
Fatty acid pathway				
PGSC0003DMT400034740	Acyl-CoA--sterol O-acyltransferase 1	T156M	0.002	-5.6
PGSC0003DMT400034544	Wax synthase isoform 3	I5T	0.011	-3.3
PGSC0003DMT400034544	Wax synthase isoform 3	W68R	0.002	-14.0
PGSC0003DMT400034544	Wax synthase isoform 3	A241E	0.011	-3.8
PGSC0003DMT400034542	Wax synthase isoform 3	Y22H	0.017	-4.6

Table 3.10: Statistics of protein 3D structural predictions of native and variant alleles in F06037 and Russet Burbank. Results for 3D protein structure prediction by I-TASSER and superposed structures of 3D proteins by UCF chimera). (C score: Confidence score, TM score: Template Modeling score, RMSD: root mean square deviation, SDM score: Structural Distance Measure, Q-score: Quality score).

S.No	Native allele (F06037)	I-TASSER results			UCF chimera superposition results		
	Mutant allele (Russet Burbank)	c score	Estimated TM score	Estimated RMSD score	RMSD between the structures	SDM	Q-score
1	CCR^{S54}, CCR^{W75}	0.83	0.83±0.08	3.8±2.6Å			
	CCR^{S54F}, CCR^{W75R}	0.77	0.82±0.09	3.9±2.6Å	0.747	14.873	0.942
2	CCR^{S205}, CCR^{W226}	-1.42	0.54±0.15	9.8±4.6Å			
	CCR^{S205F}, CCR^{W226R}	-1.49	0.53±0.15	10.0±4.6Å	0.995	20.265	0.743
3	CCR^{A79}	1.52	0.93±0.06	3.3±2.3Å			
	CCR^{A79T}	1.49	0.92±0.06	0.92±0.06	0.257	5.139	0.993
4	$PAL1^{A3, C535, I489}$	2	0.99+/-0.04	3.5+/-2.4			
	PAL1 A3S, C535Y, I489V	2	0.99+/-0.04	3.5+/-2.4	0.178	3.565	0.996
5	$4CL1^{Y54, V189, G383}$ (a)	0.2	0.74±0.11	6.8±4.0Å			
	$4CL1^{Y54H, V189M, G383D}$ (a)	0.5	0.78±0.10	6.1±3.8Å	0.487	9.715	0.97
6	$4CL1^{Y54, V189, G383}$ (b)	1.76	0.96±0.05	3.8±2.6Å			
	$4CL1^{Y54H, V189M, G383D}$ (b)	1.07	0.86±0.07	5.2±3.3Å	0.46	9.184	0.97

CCR - cinnamoyl-CoA reductase, PAL - phenylalanine ammonia-lyase, 4CL- 4-coumarate: CoA ligase. Native alleles are in non-italics and mutant alleles are given in italics. An amino acid, positions, and substitution changes are superscripted. S- Serine, W- Tryptophan, R- Arginine, A- Alanine, T- Threonine, C- Cysteine, I- Isoleucine, V- Valine, Y- Tyrosine, G- Glycine, D- Aspartic acid, H- Histidine. (a) Allele of PGSC0003DMT400037485 (4CL1) transcript (b) Allele of transcript PGSC0003DMT400037490 (4CL1).

Figure 3.1: Classes of SNPs (transition and transversion) detected in Russet Burbank compare to two late blight resistance potato cultivars.

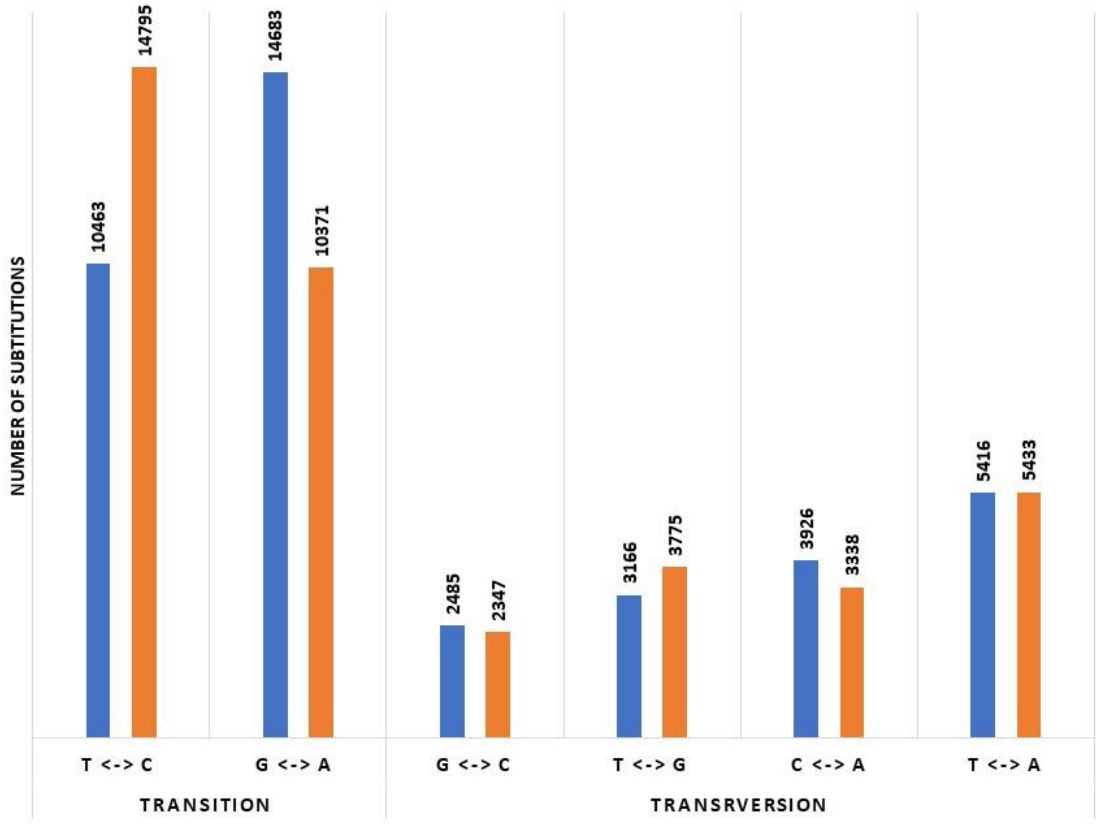


Figure 3.2 (a). Impact categories of SNPs specific to Russet Burbank potato cultivar. (b). Impact categories of InDels specific to Russet Burbank**.**

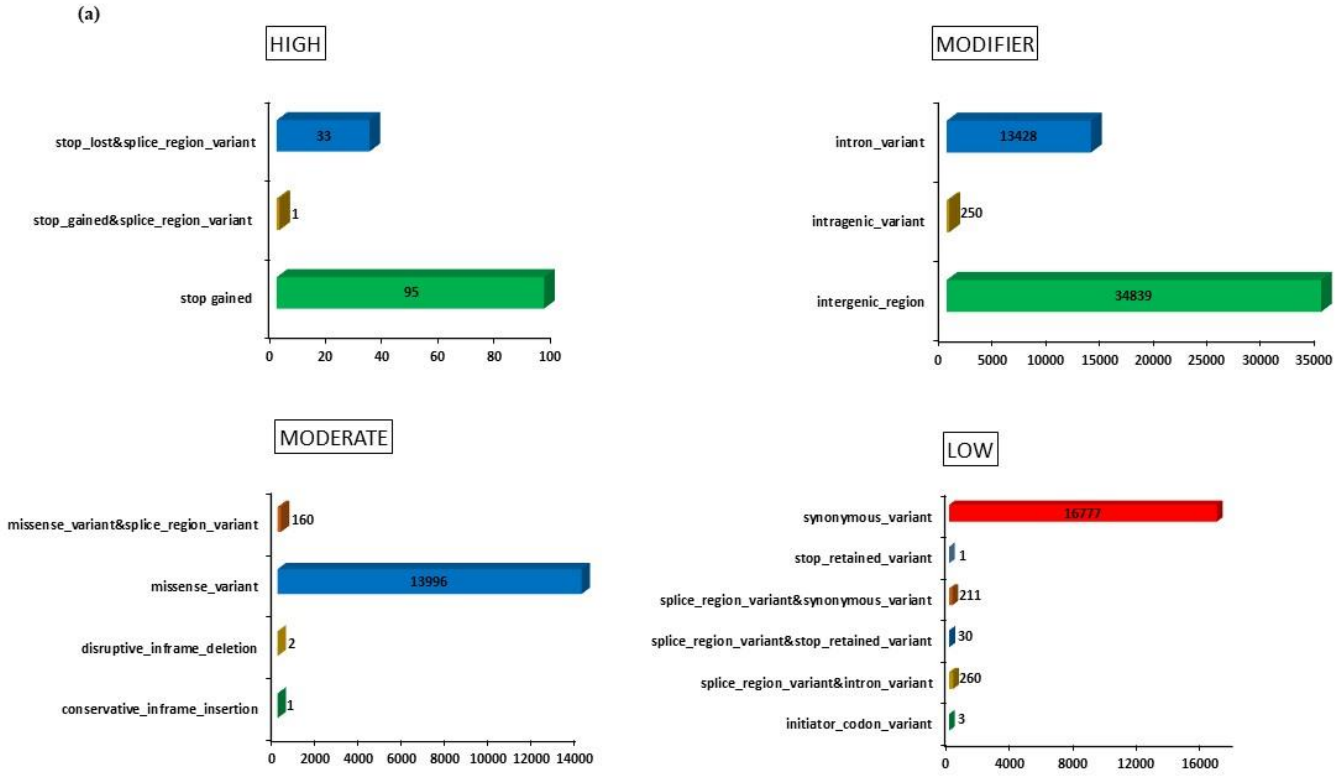


Figure 3.2: (a). Impact categories of SNPs specific to Russet Burbank potato cultivar. (b). Impact categories of InDels specific to Russet Burbank**.**

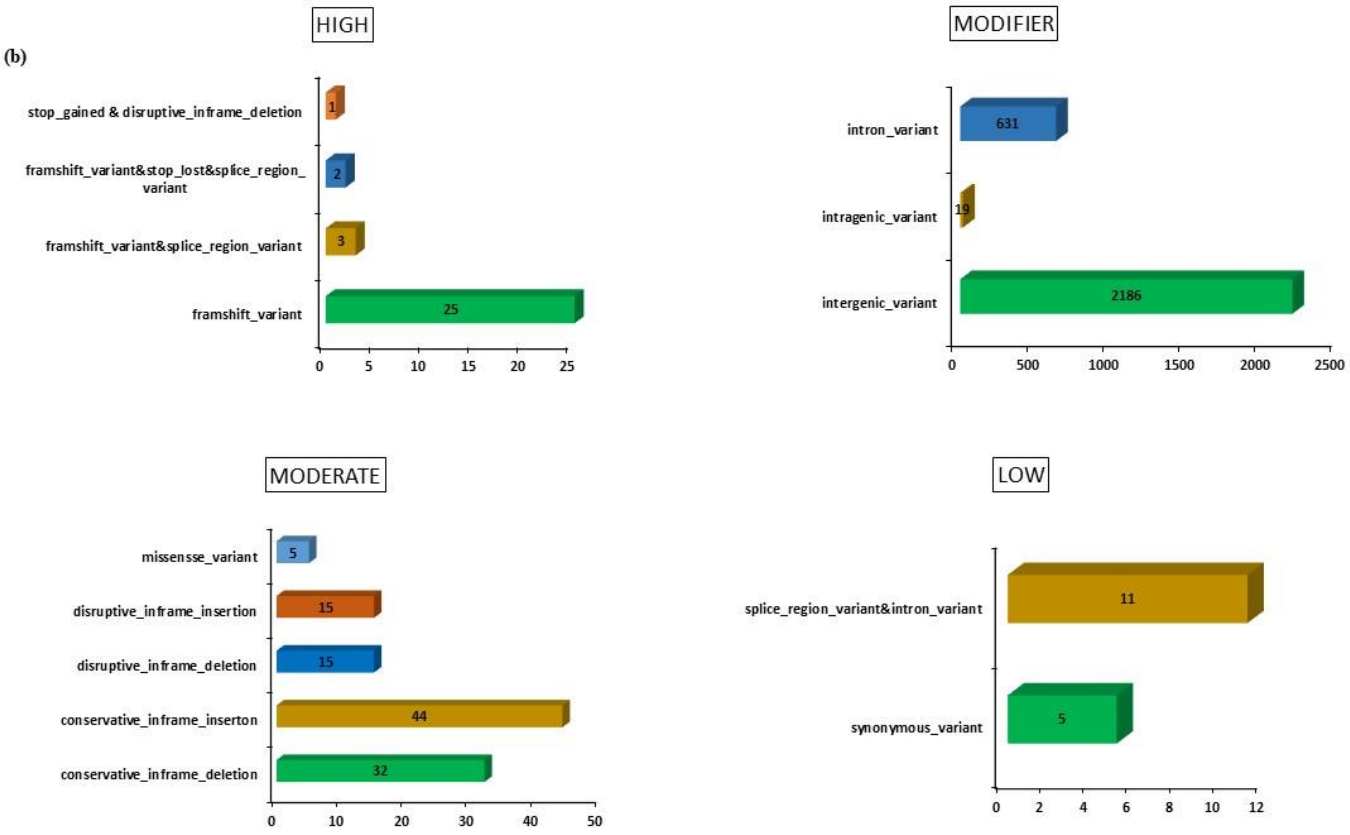


Figure 3.3: Flow chart diagram for SNPs/InDels identification from RNA-seq of late blight resistant (F06025 and F06037) and susceptible cultivars (Russet Burbank); stepwise annotation procedure to predict deleterious mutations and their impact on the structure.

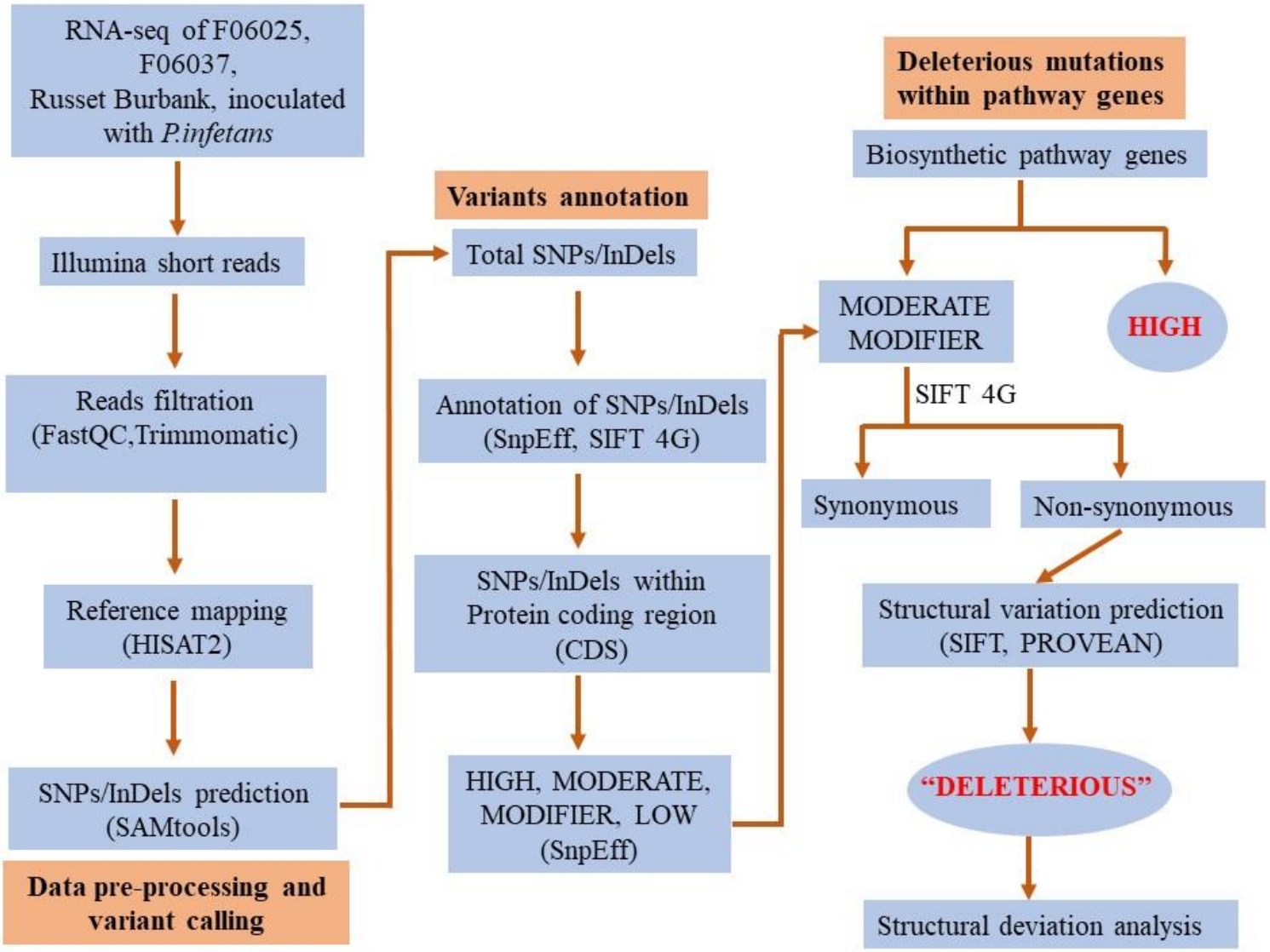


Figure 3.4: Density distribution of (a) SNPs; (b) InDels discovered within biosynthetic pathway genes detected across the 12 chromosomes of the Russet Burbank genotype. A) All the pathway genes distributed across the reference genome. SNPs or InDels of Russet Burbank belong to the phenylpropanoid pathway (B), flavonoid pathway (C), and fatty acid pathway genes or transcripts (D).

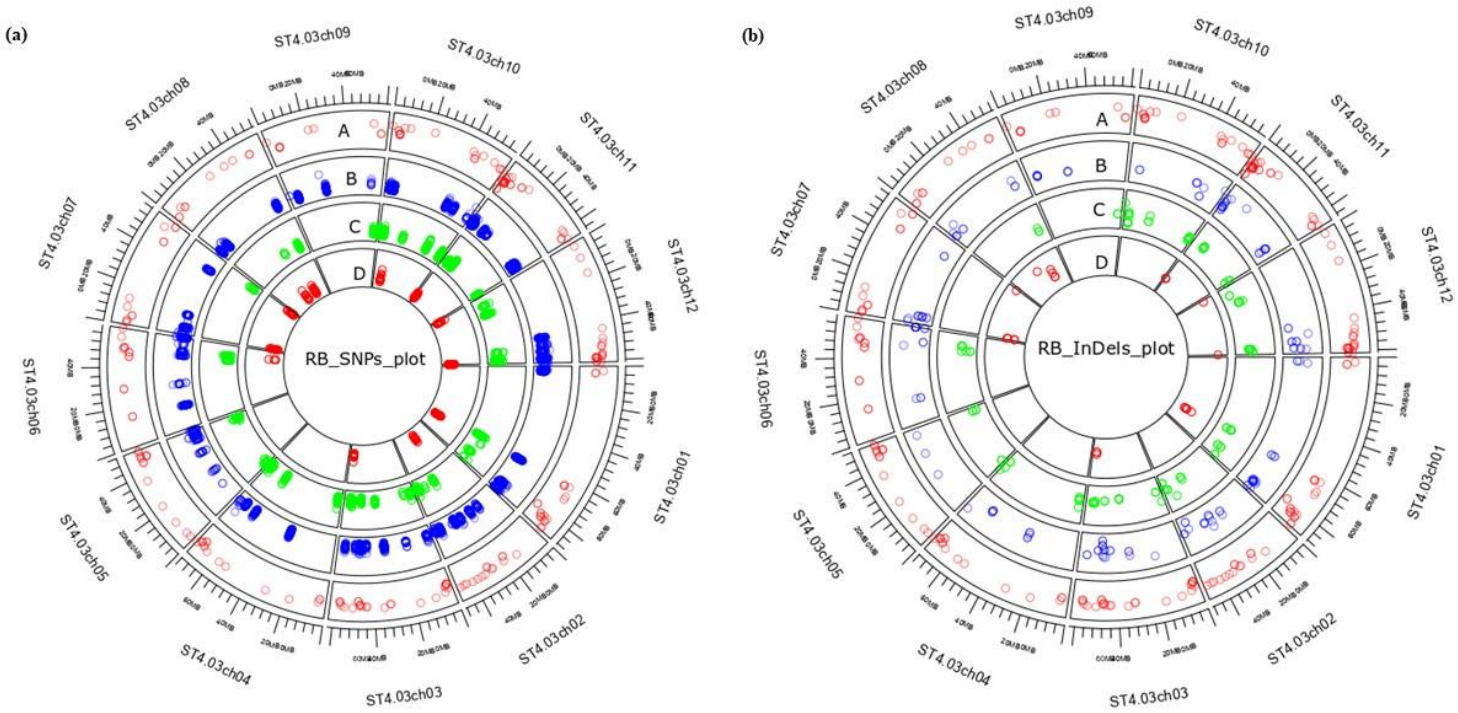
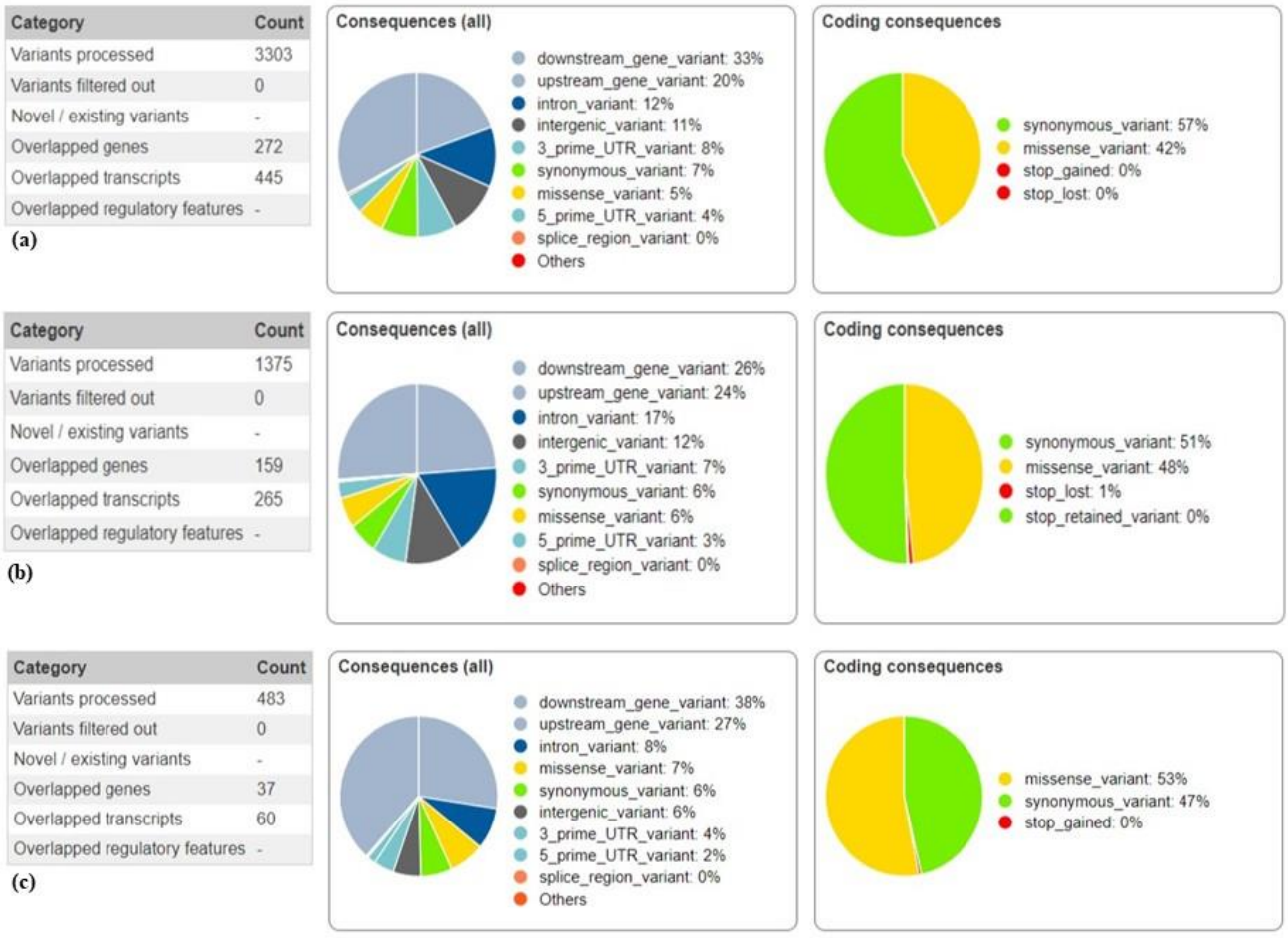


Figure 3.5: Different types of Russet Burbank specific SNP mutations identified in (a) phenylpropanoid biosynthesis; (b) flavonoid pathway; (c) fatty acid biosynthesis pathway genes.



Category	Count
Variants processed	3303
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	272
Overlapped transcripts	445
Overlapped regulatory features	-

(a)

Category	Count
Variants processed	1375
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	159
Overlapped transcripts	265
Overlapped regulatory features	-

(b)

Category	Count
Variants processed	483
Variants filtered out	0
Novel / existing variants	-
Overlapped genes	37
Overlapped transcripts	60
Overlapped regulatory features	-

(c)

Figure 3.6: qRT-PCR expression studies of 10 transcripts (Table 3.2) in Russet Burbank and F06037 following *Phytophthora infestans* inoculation, at 2 days post inoculation (2 dpi). Transcripts with HIGH impact mutations identified in Russet Burbank were selected. (a). PGSC0003DMT400029815; (b). PGSC0003DMT400006135; (c). PGSC0003DMT400039064; (d). PGSC0003DMT400007990; (e). PGSC0003DMT400076179; (f). PGSC0003DMT400084273; (g). PGSC0003DMT400007961; (h). PGSC0003DMT400074657; (i). PGSC0003DMT400034544. Comparative expression of transcripts: a-f) at 48 hours in mock (no pathogen; control) and pathogen inoculated Russet Burbank and F06037 normalized to the expression data of *StEF1αI* and *b-tubulin* reference genes. Significant differences in expression levels of resistant genotypes compared with susceptible genotype using Student's t-test are indicated: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.

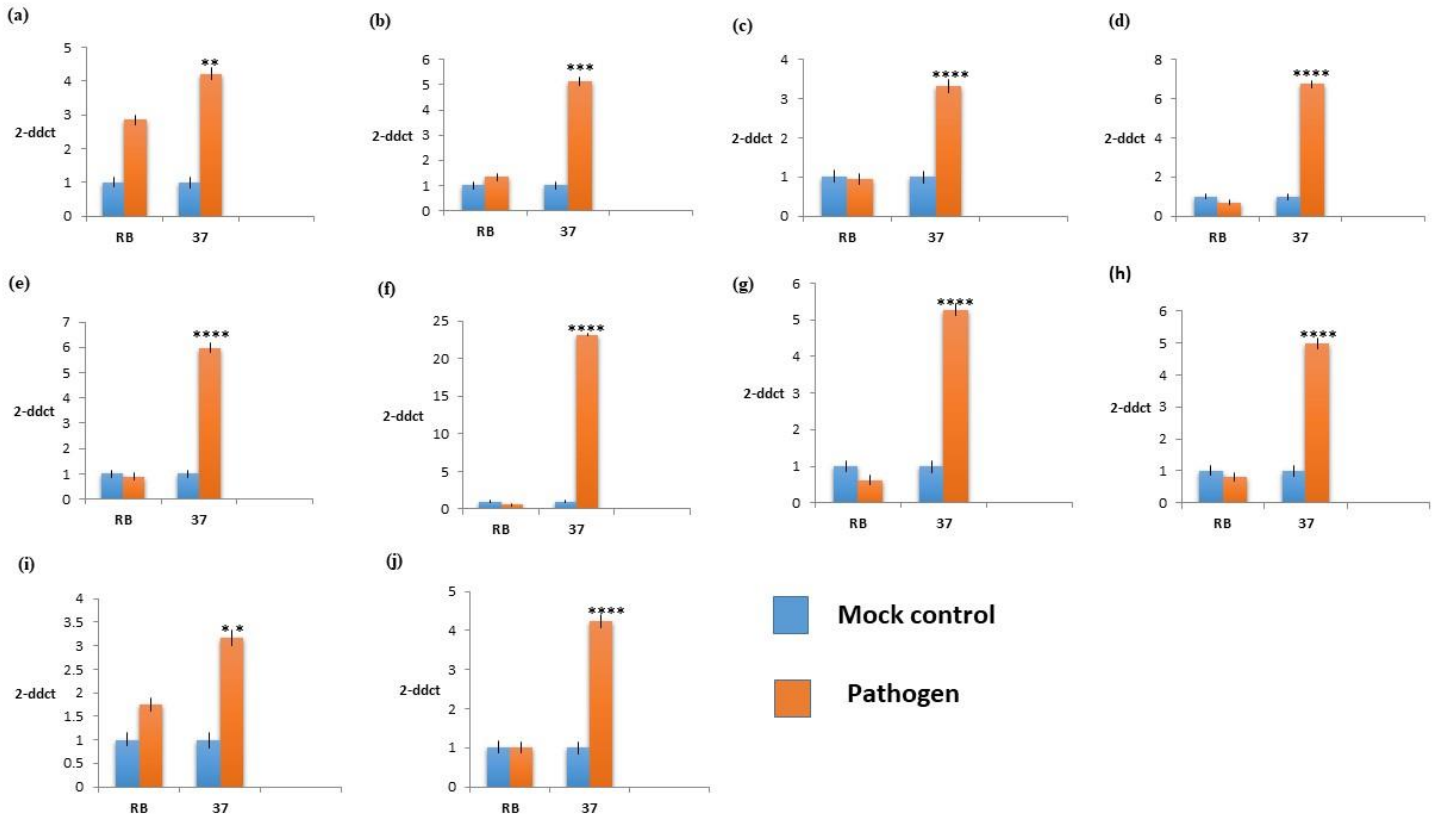


Figure 3.7: 3D structure of proteins in Russet Burbank and F06037 as predicted by I-TASSER, and structural superposition of protein through Chimera. Protein structure disorder for wild type and variant allele (native/mutant alleles): (a) PGSC0003DMT400001400 (CCR$^{S54,\ W75}$/ *CCR$^{S54F,\ W75R}$*); (b) PGSC0003DMT400001401 (*CCR$^{S205F,\ W226R}$*/ CCR$^{S205,\ W226}$); (c). PGSC0003DMT400065273 (CCR^{A79}/ *CCR^{A79T}*); (d) PGSC0003DMT400049886 (PAL1$^{A3,\ C535,\ I489}$/*PAL1 $^{A3S,\ C535Y,\ I489V}$*); (e) PGSC0003DMT400037485 (4CL1$^{Y54,\ V189,\ G383}$/ *4CL1$^{Y54H,\ V189M,\ G383D}$*); and (f) PGSC0003DMT400037490 (4CL1$^{Y54,\ V189,\ G383}$/ *4CL1$^{Y54H,\ V189M,\ G383D}$*). Root-mean-square deviation (RMSD) values are given (Table 5) for each 3D structure and superposed protein 3D structures (Table 5). All the mutations were categorised as a deleterious nonsynonymous/missense mutation based on SIFT prediction and PROVEAN scores.

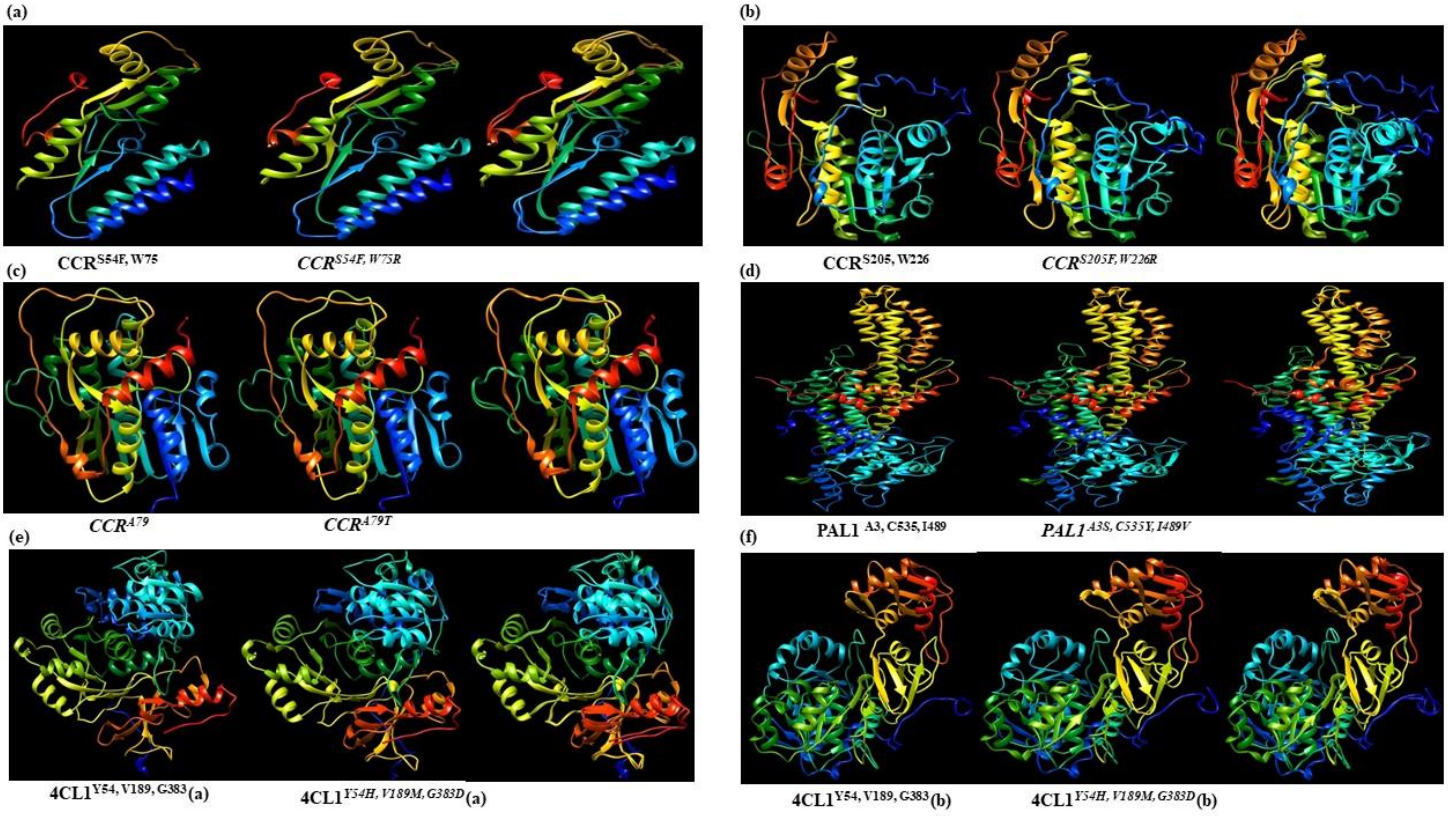
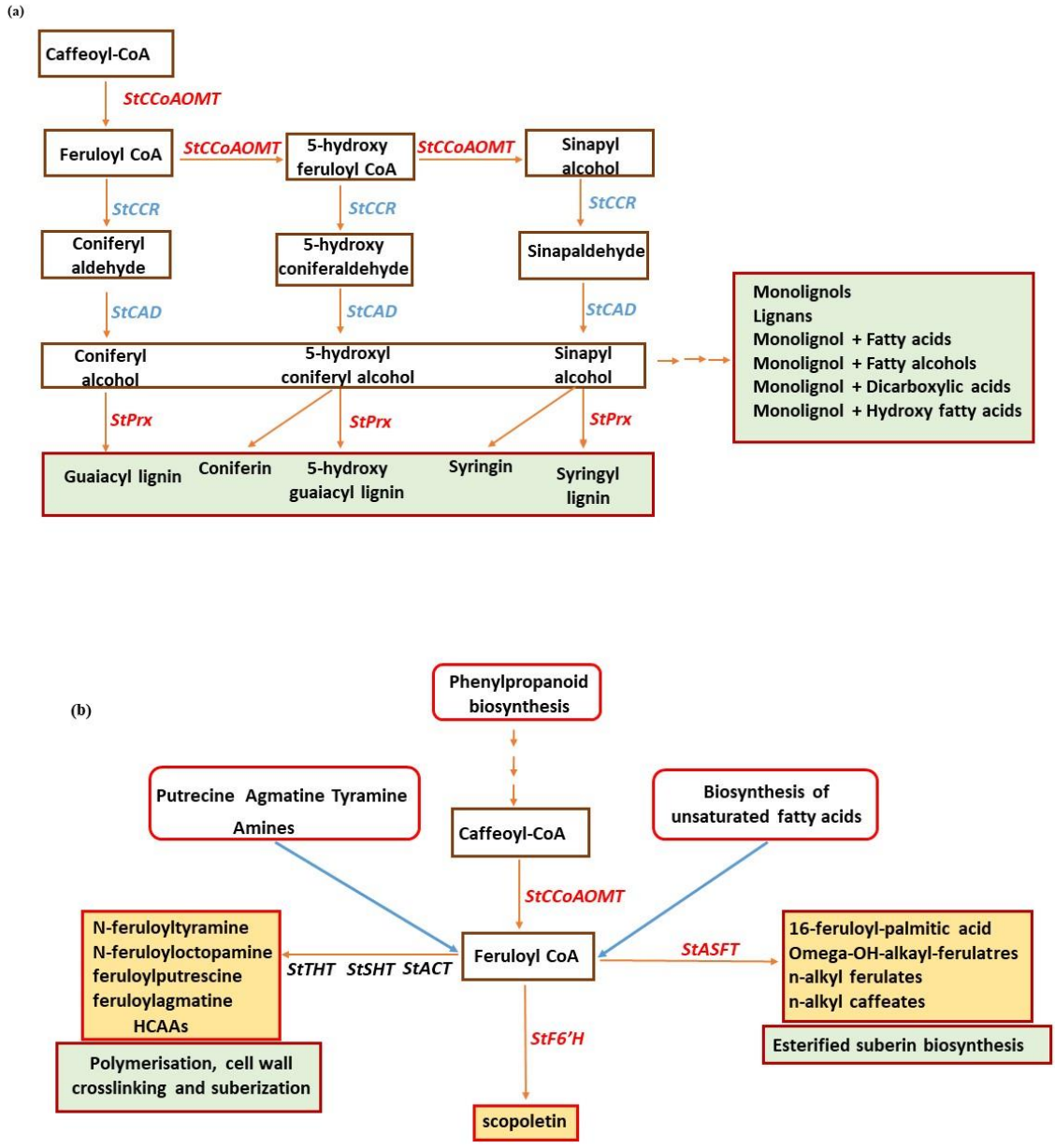


Figure 3.8: Plausible late blight resistance mechanisms, by lignification and suberization (secondary cell wall reinforcement). Pathogen (*Phytophthora infestans*) induced pathways leading to cell wall reinforcement. Genes and respective transcripts with HIGH impact mutations are in red; genes with non-synonymous (missense) are in blue. (a) Lignin and monolignol biosynthesis; (b) HCAAs, aliphatic and aromatic suberin synthesis and scopoletin biosynthesis.



CONNECTING STATEMENT FOR CHAPTER IV

Chapter III, entitled “Identification and functional characterization of late blight resistance polymorphic genes in Russet Burbank potato cultivar” Running title: Disease resistance genes polymorphic in Russet Burbank potato authored by Niranjan Hegde, Dadakhalandar Doddamani, Ajjamada C. Kushalappa has been accepted for publication in Functional Plant Biology. Co-author's contribution towards the manuscript is provided in the “contributions of authors” section. The entire thesis was formatted to one type which is consistent throughout the thesis.

In chapter III, the RNA-seq data of two resistant genotypes (F02037 and F02025) and one susceptible cultivar (RB), inoculated with *P. infestans* sporangia was reanalyzed to detect single-nucleotide polymorphisms (SNPs) and small insertion and deletions (InDels) in Russet Burbank cultivar, as compared with resistant F06037 and F06025. After the identification of SNP and InDel validations, the polymorphic genes were subsequently mapped to metabolic pathways, with a major focus on RR metabolites. The previous studies on different crops such as wheat, barley and potato inoculated with pathogens have revealed the role of phenylpropanoids conjugated metabolites in disease resistance (Yogendra et al., 2015b; Kumar et al., 2016a; Dhokane et al., 2016). The phenylpropanoid pathway intermediate such as hydroxycinnamic acid CoA-thioesters has been proved to be the substrates for the biosynthesis of complex conjugated hydroxycinnamic acid amides, which are known to be involved in secondary cell wall thickening. Besides, several studies have proved that the phenylpropanoid pathway intermediate metabolites serve as substrates for the biosynthesis of several complex polymers and conjugated metabolites, such as lignin, flavonoids, isoflavonoids, coumarins, wall-bound phenolics, suberin and fatty acids. All these complex compounds and polymers cannot be easily broken down by enzymes secreted by pathogens. Finally, only one important polymorphic gene with detrimental mutations in Russet Burbank was selected as the candidate gene for further genome editing.

In chapter IV, the gene (two transcript variants) encoding caffeoyl-CoA O-methyltransferase which showed a deleterious mutation in Russet Burbank was knocked-out in a resistant genotype F06037, using CRISPR-Cas9 to unravel the late blight resistance functions of this gene. Furthermore, the mutation in Russet Burbank cultivar was corrected by CRISPR-Cas9 based gene

targeting. The resistance effects of this gene transcripts were validated based on disease severity assessment and pathogen biomass quantification, and the resistance mechanism of the functional gene was confirmed based on metabolite profiling of the edited plants relative to control.

Also, Chapter IV entitled “The caffeoyl-CoA O-methyltransferase gene SNP replacement in Russet Burbank potato variety enhances late blight resistance through cell wall reinforcement” Niranjan Hegde, Sripad Joshi, Nancy Soni and Ajjamada C. Kushalappa was submitted to the peer-reviewed scientific journal for publication recently (under review).

CHAPTER IV

The caffeoyl-CoA O-methyltransferase gene SNP replacement in Russet Burbank potato variety enhances late blight resistance through cell wall reinforcement

Niranjan Hegde, Sripad Joshi, Nancy Soni and Ajjamada C. Kushalappa

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

4.1 Abstract

Late blight of potato is a devastating disease worldwide and requires weekly applications of fungicides to manage. Genetic improvement is the best option, but the self-incompatibility makes potato breeding very challenging. Receptor gene replacement has increased resistance, but its durability is limited. A precise Single Nucleotide Polymorphism (SNP) mutation correction of the *StCCoAOMT* gene in Russet Burbank potato, using CRISPR-Cas9 mediated homology-directed repair, significantly enhanced (Fold Change = 21.14; $P<0.001$) resistance to late blight and as well decreased the pathogen biomass. Metabolic profiling identified significant increases in the accumulation of resistance-related metabolites involved in suberization and lignification. In addition, in a late blight resistant genotype (F06037), the knock-out of two transcript variants of the *StCCoAOMT* gene confirmed the late blight resistance function. This gene, if mutated, can be replaced in any potato cultivar to enhance late blight resistance, provided it is associated with the rest of the gene hierarchy to biosynthesize the resistance metabolites.

Keywords:

Potato late blight, Caffeoyl-CoA-methyltransferase, Feruloyl CoA, CRISPR-Cas9, genome editing, homology-directed repair (HDR), geminiviral replicons, feruloylated polysaccharides, secondary cell wall reinforcement.

4.2 Introduction

Diseases are one of the major constraints in plant agriculture. Breeding aimed at improving various traits have some disease resistance, but a few pathogens continue to the plaque with high economic losses (Fry, 2008). The late blight of potato caused by *Phytophthora infestans*, resulted in Irish Famine (1845) is still a problem requiring weekly applications of fungicides to manage it under commercial conditions (Haverkort et al., 2009; Arora et al., 2014). Resistance in plants against pathogens is defined as reduced susceptibility. Following pathogen perception plant membrane-localized receptor genes trigger downstream hierarchies of regulatory R genes and resistance metabolite biosynthetic R genes, in the metabolic pathway network, and as well the R genes that code for resistance proteins (Kushalappa et al., 2016a). These metabolites are antimicrobial phytoalexins that suppress pathogen progress in plants or as polymers, the conjugated forms of which are deposited to reinforce the secondary cell walls to contain the pathogen to initial infection area (Yogendra et al., 2015; Kumar et al., 2016a).

Improvement of resistance against pathogen attack has been mainly focused on receptor genes. Plants have thousands of receptor genes to detect thousands of non-specific elicitors and specific effectors produced by pathogens. Pyramiding of some of these R genes failed to confer durable disease resistance, over time, space and environment (Uhse and Djamei 2018; van Esse et al., 2020). Although these R genes are surveillance genes, they are very important to perceive the pathogen and trigger the downstream resistance R genes at an early stage of pathogen invasion. All the same, some of these receptor R genes are associated with durable resistance, and generally these R genes induce downstream resistance R genes that biosynthesize cell-wall bound polymers to contain the pathogen (Andersen et al., 2018; Nelson et al., 2018). The comprehensive knowledge on the hierarchies of R genes, regulatory and biosynthetic, involved in a plant to resist a pathogen is still elusive.

Plants possess several mechanisms of resistance. The total resistance in a plant is due to the cumulative resistance effects of several resistance metabolites and proteins. However, a commercial plant need not biosynthesize all the resistance metabolites known in plants; a few important metabolites or proteins can offer high levels of resistance under field conditions. A commercial cultivar that is susceptible to disease also has most of the R genes to resist pathogens like a resistant genotype, but a few may be mutated, which disable them to biosynthesize a given metabolite(s) rendering it susceptible. These mutations occur due to hybridization, radiation and gene transfer. If the R genes that biosynthesize a set of specifically required resistance metabolites are polymorphic or mutated, they can be replaced with functional R gene segments from a resistant genotype based on genome editing. The stacking of a few functional R genes should confer the required level of multiple disease resistance (MDR) under commercial conditions. Although MDR has been successfully used in crops the molecular mechanisms controlling MDR are largely unknown (Wiesner-Hanks and Nelson 2016).

CRISPR-cas9 genome editing has been successfully used to improve the disease resistance in several plants by knocking-down the susceptibility genes (S-genes) (Wang et al., 2014; Peng et al., 2017; Fister et al., 2018). All these studies were successful in gene knock-out using CRISPR-Cas9, which technically relies on a non-homologous end joining (NHEJ) dependent repair system (Puchta et al., 1996; Iliakis et al., 2004; Symington and Gautier 2011). After the CRISPR-Cas9 induces a double-strand break or a nick, plant cells either chose NHEJ or homology-directed repair (HDR) pathway to repair on its own. Normally, the NHEJ pathway is frequent in plants and introduces InDels (insertion and deletion) at the targeted nick site on the genome (Collonnier et al., 2017). On the other hand, HDR is dependent on the availability of a repair template during the natural repair mechanism (Puchta et al., 1996; Symington and Gautier 2011; Steinert et al., 2016). The HDR based repair has proven to be inefficient for the practical implications due to the lack of supply of the HDR repair template (Collonnier et al., 2017; Hahn et al, 2018; Mao et al., 2019). However, there are a few studies, reported harnessing HDR to carry out gene targeting (gene replacement or knock-in) in plants (Zhao et al., 2016; Hahn et al., 2018; Miki et al., 2018; Li et al., 2018). Gene targeting can be made available for fixing different types of mutations including SNPs and InDels within the plant genome. Among all the available methods, geminiviral replicon based CRISPR-Cas9 system is found to be very efficient in carrying out gene targeting, wherein it

is possible to produce abundant HDR repair template which can be made available during the double-stranded break repair (Wang et al., 2017; Dahan-Meir et al., 2018).

A previous analysis of RNA-seq data from Russet Burbank cultivar (susceptible to late blight) enabled the identification of point mutations affecting two transcript variants of the *StCCoAOMT* gene based on molecular functions (Hegde et al., 2020). Here, in this study, the knock-out of two transcript variants of the *StCCoAOMT* gene (not affecting the functions of the other two variants of the same gene) in a resistance genotype (F06037) confirmed the late blight resistance function. Furthermore, the correction of point nonsense mutation within these transcript variants of *StCCoAOMT* in a susceptible cultivar (Russet Burbank) conferred enhanced late blight resistance. Metabolic profiling of edited and non-edited plants, inoculated with mock or pathogen, revealed the involvement of the *StCCoAOMT* gene in encoding caffeoyl-CoA O-methyltransferases which convert caffeoyl-CoA to feruloyl-CoA to produce feruloylated polysaccharide and ferulic acid polymers, upon pathogen challenge, to reinforce the cell wall.

4.3 Materials and methods

4.3.1 Phylogenetic analysis

All sequences in the study were obtained from NCBI database. Phylogenetic analysis was carried out using MEGA 7.0 by the maximum likelihood method based on the poisson correction model (Kumar et al., 2016b). The analysis involved 22 amino acid sequences from different plant species. A bootstrap test was done with 1000 replicates. Besides, the exon and intron structure is visualized based on RNA-seq data of Russet Burbank uploaded into the Phytazome JB browser (https://phytozome.jgi.doe.gov/pz/portal.html#!info?alias=Org_Stuberosum) and schematically represented later (Fig. 4.1).

4.3.2 Plant production and tissue culture

A potato late blight resistant genotype (F06037) and susceptible genotype (Russet Burbank) was obtained from Mrs. Agnes Murphy (Potato Research Centre, Agriculture and Agri-Food Canada, New Brunswick, Canada). Sprouted tubers were planted in a pot containing promix BX® and perlite. Pots were later maintained under greenhouse condition at 20±3 °C temperature and a 16h

photoperiod (intensity of ~1,500 µmol/m2/s) with (70±10%) relative humidity. Besides, for *Agrobacterium*-mediated plant transformation and regeneration, internodal segments were collected from both F06037 and Russet Burbank potato genotypes. Plantlets were grown and maintained on half-strength M516 medium (PhytoTechnology, Shawnee Mission, KS) with 3% sucrose, and 2 g/L Phytagel^{TM} (Millipore Sigma, Massachusetts, US) (Duan et al. 2012). Plantlets were later maintained in a Percival growth chamber under the condition at 22 ℃ (16 h light). Internodes collected from sterile plants were later used for the plant transformation.

4.3.3 Gene editing based on CRISPR-Cas9: vector construction

Two Single guide RNAs or dual sgRNAs (sgRNA1 and sgRNA2) were designed to target the *StCCoAOMT* gene (International Potato Genome Sequencing Consortium (PGSC ID): PGSC0003DMG400002387). Importantly, sgRNAs were designed to target the first exons of transcript variants (PGSC0003DMT400006134 and PGSC0003DMT400006135). The sgRNA expression cassettes were synthesized separately on gBlocks (Integrated DNA Technologies). sgRNA1 expression cassette with potato U6 promoters and sgRNA2 with *Arabidopsis* U6 promoter were cloned into the pUC119-MCS plasmid using *HindII/KpnI* and *PstI/SacI* as described before (Li et al., 2013) (Fig. 4.2a & Appendix 4.1). Multiple guide RNA expression cassettes were later cloned into pFGC-pco-Cas9 (Li et al., 2013). Also, the repair template (HDR template) sequences were synthesized on gBlocks (Integrated DNA technologies). The repair template was cloned into pLSL.R plasmid using *Xho I* restriction site between large intergenic region (LIR), and short intergenic region (SIR) (Fig. 4.2b).

4.3.4 Plant transformation and protoplast transformation of potato and genotyping of edited plants

Agrobacterium-mediated transformation of internodes (susceptible genotype: Russet Burbank) was performed without any modifications, as previously described (Duan et al., 2012). Similarly, for the resistance genotype (F06037), the different regeneration media and methods have been used (Kumlay and Ercisli 2015). In both cases, the plantlets were regenerated from the callus and selected on 2 mg/L of phosphinothricin (L-PPT, glufosinate-ammonium) (Fig. 4.3). Plantlets regenerated were maintained to produce T0 and first clonal generation (CG1) respectively, as

explained in plant production and tissue culture method section. The CRISPR-Cas9 vector constructs were used for the transfection of both protoplast and internodes; only the latter was regenerated into plants. Complete protoplast isolation and PEG-mediated potato protoplast transformation were followed as reported earlier (Clasen et al., 2015). Genomic DNA was isolated from protoplast (at 48hrs after transformation), T0 and progenies of T0 plants or first clonal generation (CG1). *StCCoAOM*T gene region (950bp) covering the target site was amplified using primers *CCoAOMT*_Forward: GTGAACGCTTTCCTCTGGCA and *CCoAOMT*_Reverse: AGGTAGGAGTTTGGAGGAAGTCT and Phusion polymerase (Invitrogen, Carlsbad, USA). The amplified gene segment was further cloned into pGEM®-T Easy (Promega Corporation, Madison, Wisconsin, United States) by TA cloning to determine the allelic composition (monoallelic, biallelic, homozygous and heterozygous) of CG1 plants. A minimum of 12 colonies was randomly picked, the plasmid was isolated (Qiagen, Hilden, Germany) and sequenced as genotyping followed in other studies (Andersson et al., 2017). InDel (insertion/deletions) variations within the target sequences were later confirmed by Sanger sequencing and ICE v2 CRISPR analysis tool (ICE v2) from Synthego (https://ice.synthego.com/#/).

4.3.5 Transcript variants expression based on reverse transcription PCR (RT-PCR)

The gene expression studies for the transcript variants of the *StCCoAOMT* gene (PGSC0003DMT400006134 and PGSC0003DMT400006135) were performed using transcript specific RT-PCR primers (Table S1). The RNA extracted from leaves of pathogen inoculated Russet Burbank and Russet Burbank edited *StCCoAOMT*^{KI} genotypes at 2 dpi (48 h) following the inoculation with *P. infestans* or with water (as mock control) was used for transcriptome analysis and cDNA synthesis. The first strand was synthesized using AffinityScript QPCR cDNA Synthesis Kit (Agilent Technologies). qRT-PCR was performed in a reaction volume of 10 µl consisting of 20 ng cDNA, 2 pmole of each primer, Qi-SYBR Green supermix (BioRad). The reactions were carried out in the CFX384TM Real-Time system (BioRad, ON, Canada). The results were analyzed based on delta-delta C_T (cycle threshold) method ($2^{-\Delta\Delta CT}$) (Livak and Schmittgen 2001).

4.3.6 Disease severity and pathogen biomass assessment

The experiment was conducted as a randomized complete block design with four genotypes (F06037, F06037 knock-out (*StCCoAOMT*^{KO}), Russet Burbank and Russet Burbank edited (*StCCoAOMT*^{KI}) and two inoculations (pathogen and mock), three replicates over time. The plants were inoculated with a sporangial suspension of *P. infestans*. Each experimental unit consisted of five pots, with two plants and ten leaves inoculated on the lower surface. Following inoculation, the plants were covered with plastic bags for 3 days (72 h). The relative biomass of *P. infestans* in the infected samples was quantified based on quantitative PCR (qPCR), to determine the growth of *P. infestans* on infected potato leave (Asai et al. 2008). Genomic DNA was isolated from *P. infestans* infected leaves (6 dpi) (Fig. 4.4) using a Dneasy Plant Mini Kit (Qiagen, Canada). qPCR was performed using IQ SYBR Green Supermix (Bio-Rad, Canada) in a CFX384TM Real-Time System (Bio-Rad, Canada) according to the manufacturer's instructions, using specific primers to amplify and detect *P. infestans* DNA (O-8_Forward and Reverse) and potato DNA (*Stβ-tubulin* _Forward and *Stβ-tubulin* _Reverse; *StEf1-α*_Forward and *StEf1-α*_Reverse) (Table 4.1). Also, disease severity was quantified by measuring lesion length using a digital calliper at 3 d intervals until 9 d post-inoculation (dpi). Lesion length (mm) was used to calculate the area under the disease progress curve (AUDPC).

4.3.7 LC-high-resolution MS (LC-HRMS) analysis for semi-targeted metabolic profiling

The experiment was conducted as a completely randomized block design with two genotypes (*StCCoAOMT*^{KI}: considered here as a resistance genotype and Russet Burbank: susceptible genotype) and two inoculations (mock and five plants with a total of 10 stems inoculated for each treatment in each experimental unit). The stems inoculated with the mock solution and *P. infestans* and samples were collected at 3 dpi. The metabolites were extracted using aqueous methanol and analyzed using liquid chromatography and high-resolution mass spectrometry (LC-HRMS, 1C-Orbitrap MS, Thermo scientific). The detailed sample collection was described earlier (Yogendra et al., 2014b).

4.3.8 LC-HRMS data processing and identification of RR metabolites

The LC-HRMS output raw data was processed, peaks were annotated using mzMine software (MZmine 2 with the high sensitivity peak detection algorithm XCMS (Pluskal et al., 2010). MZmine 2 output file with observed masses and their abundance (relative intensity) were imported to MS Excel. Inconsistent peaks among the replicates and adducts before the identification of RR metabolites. Monoisotopic mass peak intensity (m/z) data were subjected to pairwise Student's t-test analysis. Mono isotopic mass with significantly higher abundances in resistant (*StCCoAOMT*^{KI}) than in susceptible (Russet Burbank) genotypes were considered as resistance-related (RR) metabolite (Kushalappa and Gunnaiah 2013). These were further termed as RR induced (RR metabolites) (RRI=[(RP/RM>1.0)/(SP/SM>1.0)], R is resistant (Russet Burbank edited *StCCoAOMT*^{KI}), S (Russet Burbank) is susceptible genotype; P is a pathogen and M is mock solution inoculated.

4.4 Results

4.4.1 *StCCoAOMT* gene in Russet Burbank is mutated and it fails to produce a functional protein

In our previous work, the RNA sequences of Russet Burbank cultivar, inoculated with *P. infestans*, identified several mutations that were classified into different types based on their impact on structure and function functions of the protein. The RNA sequencing of Russet Burbank revealed a point mutation in the *StCCoAOMT* gene (The study identified a stop gained mutations p.Glu80* (Glutamic acid (E) to STOP) and p.Glu42* (Glutamic acid (E) to STOP) within the transcripts (PGSC0003DMT400006134: *StCCoAOMT* p.Glu80* and PGSC0003DMT400006135 (a primary transcript, *StCCoAOMT* p.Glu42*) encoding *StCCoAOMT* in Russet Burbank (Fig. 4.1). Further, Sanger sequencing and gene prediction analysis (GeneScan: http://argonaute.mit.edu/GENSCAN.html) confirmed the single base change that created a premature nonsense codon, truncating the StCCoAOMT protein encoded by both the transcript variants. The resulted protein lacks 96 (PGSC0003DMT400006135) and 58 (PGSC0003DMT400006134) amino acids from N-terminus when compared to the wild type

protein. The primary transcript of *StCCoAOMT* (PGSC0003DMT400006135) consisted of 4,382 bp, encoding for 282 amino acid residues. The amino acid sequence, along with 22 amino acid sequences from different plant species, was subjected to phylogenetic analysis. It was found out that *StCCoAOMT* was highly similar to *CCoAOMT* wild type in tomato species (*Solanum pennellii* and *Solanum chilense*) with (86% identity, 94 % similarity). Also, the RNA-sequencing data revealed the exon/intron structure of the primary transcript (PGSC0003DMT400006135), which along with another variant PGSC0003DMT400006134 found to have functional protein domain (Fig. 4.1a & b). Consequently, functional annotation of the domain showed their role in suberin monomer biosynthesis.

4.4.2 *StCCoAOMT* gene knock-out in late blight resistant genotype (F06037), based on CRISPR-Cas9, validated the resistance functions of two transcripts

In the present study, plants with InDel mutations were generated targeting *StCCoAOMT* transcripts (Fig. 4.5a & b), using *Agrobacterium*-mediated CRISPR-Cas9 potato transformation of late blight resistance potato genotype (F06037). CRISPR-Cas9 transformed potato internodes successfully produced callus and regenerated into plantlets **(**Fig. 4.3). Progenies of T0 or first clonal generation (CG1) of *StCCoAOMT* knocked out plants (*StCCoAOMT*^{KO}), with deletion of -4 bps, -6bps (sgRNA1, sgRNA2) and -8bps, -2bps (sgRNA1, sgRNA2) at target sites (homozygous), were selected (Fig. 4.5). Further, the late blight resistance function of *StCCoAOMT* transcripts was characterized by quantifying disease severity and pathogen biomass in the diseased area. The homozygous knockout *StCCoAOMT*^{KO} (F06037 KO) plants showed a significant increase in late blight disease severity and pathogen biomass (Fig. 4.6a & 4.6b). Protoplast transformation assay was also conducted to prove the efficiency of CRISPR components (sgRNAs and Cas9). The use of two sgRNAs coupled with Cas9 efficiently created more than 10bps deletion/insertion mutations within the target sites and generated different mutant alleles (Fig. 4.7. 4.8, 4.9).

4.4.3 *StCCoAOMT* gene knock-in (mutation correction) in Russet Burbank, based on CRISPR-Csa9 and geminivirus-based replicons, encoded the functional transcripts

An HDR template was designed with 250 bps left and right homology from the mutation mutated protospacer adjacent motif (PAM) (MP1 and MP2) (Fig. 4.10a). The two separate plasmids were used to carry-out knock-in in the Russet Burbank genotype. Before plant transformation, the two cultures of *Agrobacterium* strain harbouring these plasmids were mixed (1:1). The first plasmid used for knock-in was the same as before used in the knock-out experiment (plasmid carrying sgRNA and Cas9). However, the second plasmid was additionally constructed to deliver the HDR template to carry out knock-in or to repair mutated alleles (*StCCoAOMT*) in the Russet Burbank genotype. We combined a few previous approaches which are reported in plants to carry out gene-knock-in or gene targeting via homologous recombination (Zhao et al., 2016; Hahn et al., 2018; Dahan-Meir et al., 2018). Moreover, the HDR template (500 bps) was designed by introducing three changes (at sgRNA1 PAM site, at sgRNA2 PAM site and one at mutation site to correct stop codon within *StCCoAOMT* alleles). As the replicons of the bean yellow dwarf virus (BeYDV) is widely used in dicot plants for increasing the gene targeting efficiency, here the HDR template was cloned in between large intergenic region (LIR), and short intergenic region (SIR) with the coding region of replication initiator protein (Rep/RepA) as previously described (Baltes et al., 2014; Butler et al., 2016). Russet Burbank plants regenerated were screened to identify HDR allele integration by performing PCR with primers targeting the genomic region away from the HDR template. The base change at the sgRNA1 PAM site also erased the restriction site (*Xho I*) which was used as a marker to identify the HDR template integration or repaired alleles (Fig. 4.10a-c). A similar strategy was used by (Li et al., 2018) to replace a gene or the allele with an SNP in one of the rice cultivars. PCR amplified products were further sequenced to identify three desired base changes within the HDR allele or repaired alleles. PCR amplified products were subjected to the restriction digestion assay to preliminarily screen for the edited Russet Burbank plants. The sequences obtained for all the putative edited F06035 and Russet Burbank genotypes were thoroughly analyzed using CRISPR Analysis Tool (ICE v2), Synthego. Four transformation events and respective plants of Russet Burbank potato plants were selected for further validation based on the ICE v2 analysis. These events clearly showed the insertion of the HDR template in all the four alleles. ICE v2 score more than 95% were considered as homozygous in case of both knock-

out and knock-in (Fig. 4.9). The homozygous insertion of HDR template (repaired allele) and CRISPR induced mutant alleles was also confirmed later by cloning and Sanger sequencing multiple times. Furthermore, the rescue of two transcript variants of *StCCoAOMT* gene (PGSC0003DMT400006134 and PGSC0003DMT400006135) by correcting the mutation (p. Glu80* and p. Glu42*) was confirmed by the expression studies. Gene expression analysis (qRT-PCR) of pathogen inoculated leaf samples (2 dpi) of the control and edited (*StCCoAOMT*^{KI}) Russet Burbank genotypes was performed. Two transcript variants: PGSC0003DMT400006134 and GSC0003DMT400006135 highly expressed in pathogen inoculated leaves of Russet Burbank edited (*StCCoAOMT*^{KI}) as compared to the control Russet Burbank genotype (Fig. 4.6c-d). Evidently, the Caffeoyl-CoA methyl transferase encoded by *CCoAOMT* gene and its isoforms have also been reported earlier in various plant species to be significantly upregulated at different levels of disease resistance (Yang et al., 2004; Draffehn et al., 2013; Yang et al., 2017) . Also, our previous study showed the high expression of the same transcript variants (above-mentioned) in resistant genotype (F06037) and no or less expression in Russet Burbank (Hegde et al., 2020).

4.4.4 The editing of *StCCoAOMT* gene in Russet Burbank significantly decreased disease severity and pathogen biomass

Disease severity was assessed in four treatments: Russet Burbank unedited control and knock-in (*StCCoAOMT*^{KI}), and F06037 control and knock-out (*StCCoAOMT*^{KO}), inoculated with pathogen. The dark brown coloured circular lesions appeared on the leaves on all the treatments at 3 dpi of the pathogen. However, the enlarged lesions appeared prominently on RB and *StCCoAOMT*^{KO} genotypes at 9 dpi. The Area Under the Disease Progress Curve (AUDPC), calculated from the lesion diameter assessed on the leaves at 3, 6, 9 dpi, was higher in Russet Burbank control (AUDPC=300.05) than in knock-in (*StCCoAOMT*^{KI} = 108.88), whereas it was lower in F06037 control (99.63) than in knock-out (*StCCoAOMT*^{KO} = 256.38) (Fig. 4.6a). Pathogen biomass in the infected leaves quantified based on qPCR at 6 dpi, the *P. infestans*-specific gene copy numbers (DNA, O-8) were significantly ($P<0.001$) higher in Russet Burbank control than in knock-in (*StCCoAOMT*^{KI}) (Fold Change = 21.14), similarly, much higher in *StCCoAOMT*^{KO} than in F06037 **(**FC =13.38) (Fig. 4.6b).

4.4.5 Resistance metabolites differentially accumulated in stems of *StCCoAOMT* edited Russet Burbank genotype

Metabolic profiling of the stems of control and edited Russet Burbank (*StCCoAOMT*^{KI}) genotype at 3 dpi following inoculation with mock and pathogen, and analysis based on LC-HRMS revealed accumulation of several resistance metabolites. We anticipated the increased accumulation of monolignols, ferulates and ferulate conjugates in the *StCCoAOMT*^{KI} genotype (Fig. 4.11). It was assumed that the replacement of mutated alleles could increase the pool of feruloyl-CoA in *StCCoAOMT* edited plants upon pathogen inoculation. Also, the mutated alleles of the *StCCoAOMT* gene were replaced in Russet Burbank thinking that the increase in the pool of feruloyl-CoA could enhance the secondary cell wall thinking compounds in response to the pathogen. Especially, the transcript variants of the gene were targeted not to alter the normal function of the other two transcript variants (PGSC0003DMT400006133 and PGSC0003DMT400006136) of *StCCoAOMT* gene. The previous studies have revelated altering of the feruloyl-CoA pool resulted in the production of lignin polymers in the plants (Li, Chen, et al., 2013; Smith et al., 2017). But, in this study, we carefully selected the transcript variants of the gene which was affected by the mutation and is known to be expressed only in response to the pathogen (Hegde et al., 2020). As expected, the metabolic profiling of pathogen inoculated leaf samples of *StCCoAOMT*^{KI} was abundant in feruloyl-CoA derived metabolites than in the unedited Russet Burbank plants, clearly supporting the previous studies (Yogendra et al., 2014; Yogendra, Kushalappa, et al., 2015). We identified several metabolites abundant in *StCCoAOMT*^{KI} plants than in control Russet Burbank with high fold changes including ferulate, 5-hydroxy-coniferaldehyde, 6-hydroxymellein (5.53), 5-O-Feruloylquinic acid (2.91), N-feruloyltyramine (2.45), 6'-hydroxyferulate, 5-hydroxyferulate (5.51), coniferyl ferulate (4.81), scopoletin (2.21) and 16-feruloyloxypalmitate (2.08). Overall, the metabolic profiling of *StCCoAOMT*^{KI} showed an elevated level of few feruloylated metabolites (Table 4.2).

Further, allelic replacement or the SNP correction (*StCCoAOMT* $^{p.\ Glu80*}$ and *StCCoAOMT* $^{p.\ Glu42*)}$ leading to two transcript variants in *StCCoAOMT*, using CRISPR-Cas9 based knock-in, in Russet Burbank resulted in a significant increase in late blight resistance. These two transcript variants of the *StCCoAOMT* gene are considered to hinder the biosynthesis of RR metabolites, mainly the

feruloyl-coA derived amide conjugates (HCCAs), fatty acids conjugates, monolignols, and scopoletin (coumarins).

This study has provided substantial evidence that editing of mutated genes in a metabolic pathway can significantly increase the disease resistance, through an increase in abundance of conjugated metabolites that deposit to enhance cell wall thickness, containing the pathogen to initial infection area, thus explaining the mechanism of resistance. It is possible that this cisgenic Russet Burbank potato cultivar should also be able to resist other pathogens. The *StCCoAOMT* gene if mutated in other commercial cultivars can be edited to enhance late blight resistance, if they are associated with the rest of the gene hierarchy to produce high amounts of conjugated metabolites to reinforce the cell wall around the pathogen invaded cells.

4.5 Discussion

Marker-assisted breeding has enabled the development of new cultivars with several desirable traits. Often breeders discard high yielding lines and as well the older commercial cultivars for lack of disease resistance. Lately, the CRISPR-Cas9 has emerged as a feasible technology to edit mutated genes in these lines to enhance disease resistance. These studies involve three major steps: i) discovery of mutated genes; ii) selection of mutated genes as potential candidates and editing to enhance disease resistance; iii) evaluation of edited plants to confirm enhanced disease resistance. The mutated genes in a potato variety can be discovered based on RNA seq and the type of mutation can be classified as detrimental if they fail to produce a protein (Hegde et al., 2020). The candidate genes for genome editing can be selected based on their impact on disease resistance, where several metabolic pathway genes have been proved to enhance resistance (Kushalappa et al., 2016b). CRISPR-Cas9 based on geminivirus replicons has emerged as a feasible tool to edit genes (Wang et al., 2017). The greenhouse and field evaluation can prove the level of enhanced resistance following the editing of a gene. This study provides a proof of concept that editing of a mutated phenylpropanoid pathway gene in a commercial potato cultivar, Russet Burbank, enhances late blight resistance.

Following pathogen invasion disease resistance genes in plants produce several transcripts and all may not produce functional proteins and enhance resistance (Wiesner-Hanks and Nelson 2016;

Andersen et al., 2018; Irani et al., 2019). Here we have identified four transcripts, of which, only one was able to produce a functional protein. The silencing of this transcript in a late blight resistant genotype, F06037, significantly decreased resistance. To confirm if a STOP-GAINED point mutation has a significant role in late blight resistance, the gene encoding *StCCoAOMT* was silenced in the resistant genotype. CRISPR-Cas9 genome editing technology has been successfully used in generating mutant alleles and stable mutant lines in different potato genotypes (Butler et al., 2015; Wang set al., 2015; Kusano et al., 2018; Nadakuduti, 2019). CRISPR-Cas9 delivery into the protoplast has been well established to carry out targeted mutagenesis and stable mutant lines with InDels in all four alleles in tetraploid potato (Andersson et al., 2017; Andersson et al., 2018). Here in this study, the *StCCoAOMT* gene was knocked out in the F06037 potato genotype based on the CRISPR-Cas9 system (Li et al., 2013). Two resistance transcripts mutated in susceptible genotype (Russet Burbank) (PGSC0003DMT400006134: *StCCoAOMT* $^{p.\ Glu80*}$ and PGSC0003DMT400006135: *StCCoAOMT* $^{p.\ Glu42*}$), affected by stop-gained mutation (induced in response to *P. infestans*), were targeted. Using multiple sgRNAs, based on sequencing, it was possible to develop stable homozygous knock-out lines (*StCCoAOMT*^{KO}), with InDels (deletion mutation of 2bps to 8bps) at the target sites. Further, two stable knock-out homozygous mutation lines (*StCCoAOMT*^{OK} CG1 plants) of F06037 were selected (Fig. 4.5b) to validate the late blight resistance functions. The mutated lines revealed the resistance function of StCCoAOMT transcript variants (PGSC0003DMT400006134 and PGSC0003DMT400006135). Allelic variation at the level of an amino acid sequence of *StCCoAOMT* resulted in varying accumulation of phenylpropanoid metabolites and programmed cell death in maize, clearly demonstrating the role of *ZmCCoAOMT2* in the biosynthesis of pathogen-induced phenylpropanoids and their further role in multiple disease resistance (Yang et al., 2017). Accordingly, the *StCCoAOMT* gene was edited to produce this transcript in Russet Burbank potato which significantly enhanced resistance. It not only reduced disease severity but also pathogen biomass. Furthermore, the metabolic profiling of Russet Burbank, control and edited, mock and pathogen inoculated plants revealed high amounts of feruloyl derivatives (Coniferyl ferulate, 5-Hydroxyferulic acid; 16-Feruloyloxypalmitate, etc.) due to enabling of *StCCoAOMT* gene to code for the functional enzyme. These feruloyl derivatives (polysaccharides, esters etc.) are known to be deposited to reinforce the cell walls to contain the pathogen to initial infection areas (Buanafina 2009; Reem et al., 2016; Zhao and Dixon 2014). The p-coumaric acids and monolignols produced in the phenylpropanoid pathway are known to

conjugate with other metabolites in phenylpropanoid, flavonoid, fatty acid, terpenoid and alkaloid pathways to produce hundreds of conjugated metabolites which reinforce cell walls in plants following pathogen invasion (Dixon et al. 2002; Kushalappa et al. 2016; Deng and Lu 2017). Some of these have been proved to confer multiple disease resistance (Wiesner-Hanks and Nelson 2016). If these genes are mutated, they can be edited to enhance multiple disease resistance. Editing of a few candidate genes in a commercial cultivar, such as Russet Burbank, can significantly enhance resistance to multiple diseases. The resistance in plants based on these conjugated metabolites should be durable as the pathogen will not be able to break this down to progress further. These metabolites are produced within 72 hpi at which stage the lesions are only less than 10 mm and thus can reduce the blights to small spots.

Lately, several genome editing tools have emerged to produce transgene-free plants, which is an important regulation in several countries. The tetraploid potatoes are in general self-incompatible and thus the segregating populations cannot be produced to screen them out. Base editing does not require double-stranded breaks and the efficiency is high, which can enable screening out the transgenes. The *StCCoAOMT* gene, if mutated in other cultivars, can be edited to enhance late blight resistance.

Table 4.1: Primers used for pathogen biomass quantification, qPCR, and PCR restriction enzyme assay (PCR-RE).

Genbank/PGSC ID name	Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
GQ371195.1	O-8	GAAAGGCATAGAAGGTAGA	TAACCGACCAAGTAGTAAA
AB061263.1	*StEf1-α*	ATTGGAAACGGATATGCTCCA	TCCTTACCTGAACGCCTGTCA
Z33382.1	*Stβ-tubulin*	ATGTTCAGGCGCAAGGCTT	TCTGCAACCGGGTCATTCAT
PGSC0003DMG400002387	*StCCoAOMT*	GACTGGCAGGCCTAGTTCAT	GTTGGCATGCGAGACACCTA
PGSC0003DMT400006134	*StCCoAOMT*	GGTTCACCCTGGGACAATAAA	CTTTGAGAGGCTCTGGTTCTC
PGSC0003DMT400006135	*StCCoAOMT*	CTCTGCTTCTCACTGCTCTTAC	CTGGCAGGCCTAGTTCATAAG

Table 4.2: Fold change in abundance of resistance-related metabolites identified in *StCCoAMT*^{KI} plants following *P. infestans* or mock solution inoculation.

ID	Observed Mass (Daltons)	Exact Mass (Daltons)	Database ID	Name	Fold Change
19594	251.0801	251.2300	5280527	N-Feruloylglycine	2.67**
20640	194.0580	194.1800	3477029	Ferulate;5-hydroxy-coniferaldehyde	5.53**
15344	368.1112	368.3000	9799386	5-O-Feruloylquinic acid;O-feruloylquinate	2.91*
1774	210.0528	209.1700	446834	5-hydroxyferulate; 5-Hydroxyferulic acid methyl ester	4.41**
7509	313.1325	313.3000	6440659	N-Feruloyltyramine	2.45**
2863	356.1267	356.4000	6441913	Cubebin; Coniferyl ferulate	4.81*
400	192.0423	192.1700	5280460	Scopoletin	2.21*
503	448.2826	447.6000	44237331	16-Feruloyloxypalmitate	2.08**
28906	774.2137	-	-	Catechin 5-O-(2-feruloyl-6-p-coumaroyl-beta-D-glucopyranoside)	8.05**

- Observed mass: One H mass (1.0078) was added to the observed mass since LC/MS analysis was carried out in a negative ionization mode.
- Fold change calculation was based on the relative intensity of metabolites: RRI=(RP/RM)/(SP/SM) (see methods for details). Significance (t-test): *P<0.05, **P<0.01.

Figure 4.1: a Exon/intron structure of *StCCoAOMT* isoforms and point nonsense mutation identified in RNA-seq data of Russet Burbank. **b** Phylogenetic analysis of StCCoAOMT (MEGA 7.0; maximum-likelihood method).

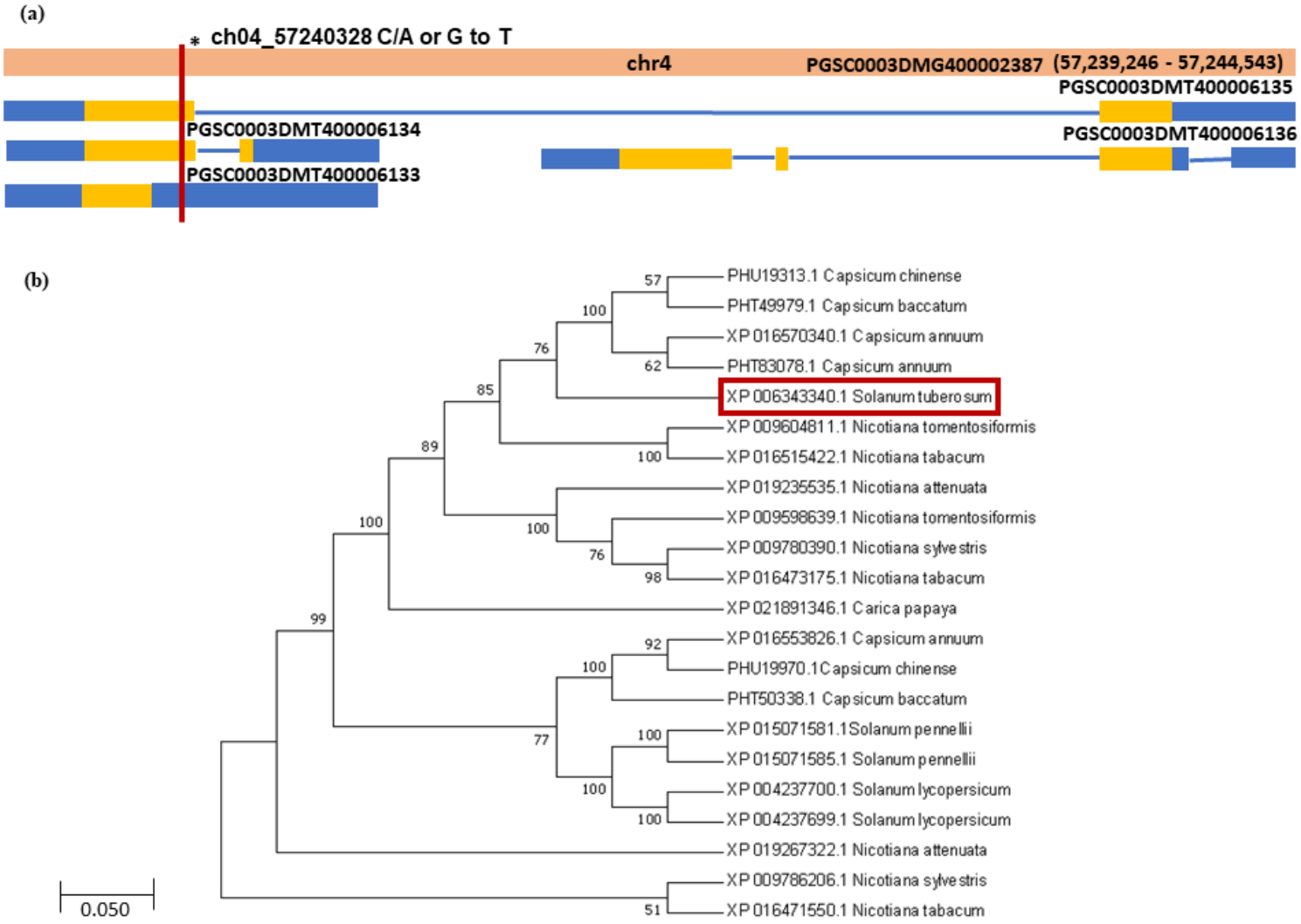


Figure 4.2: CRISPR-Cas9 vector construction and plant transformation (Li et al. 2013a). **a** Cloning sgRNAs into pFGC-pco-Cas9 transformation vector. **b** Cloning sgRNAs into pFGC-pco-Cas9 transformation vector.

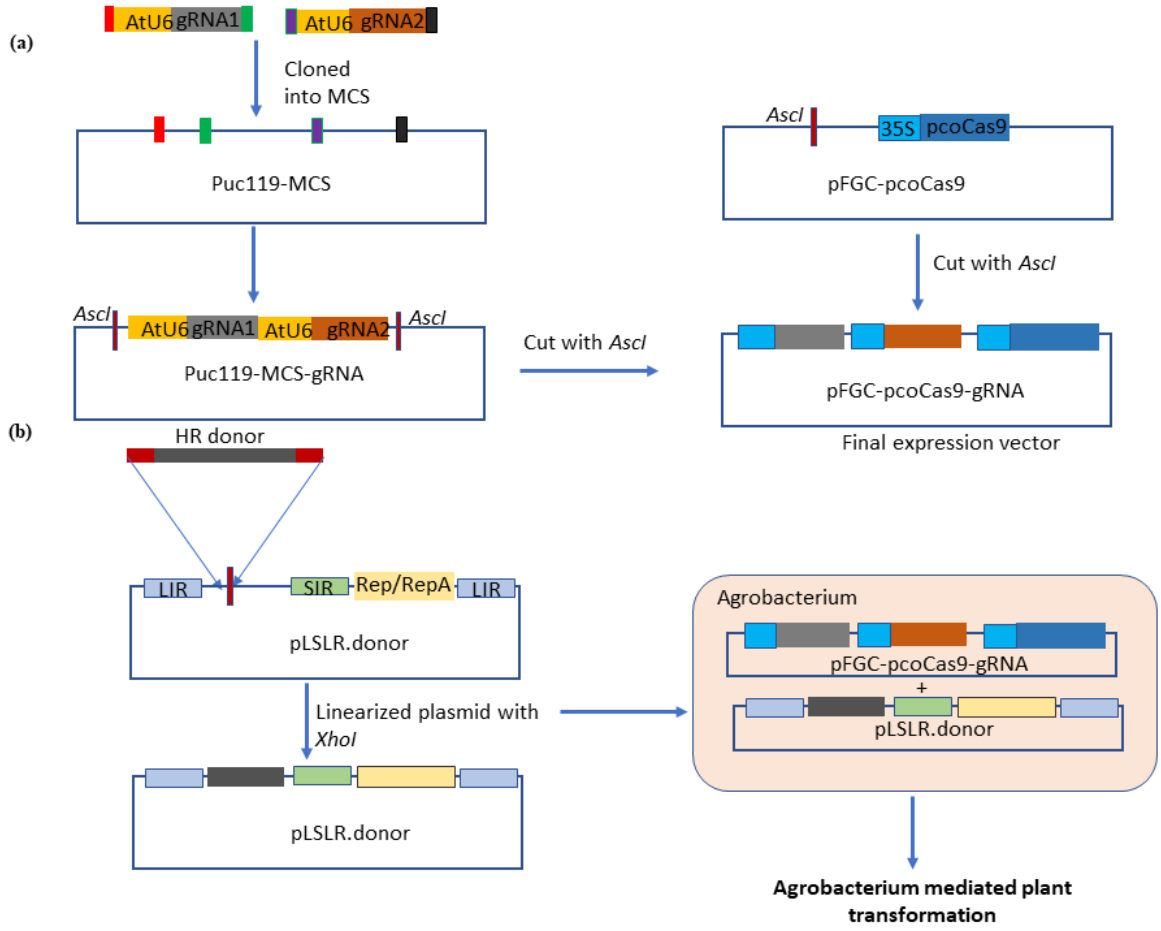


Figure 4.3: F06037 and Russet Burbank potato genotype plant regeneration from the internodal callus differentiation. *Agrobacterium*-mediated internode transformation with a plasmid carrying CRISPR-Cas9 components and regeneration (Duan et al., 2012)**.**

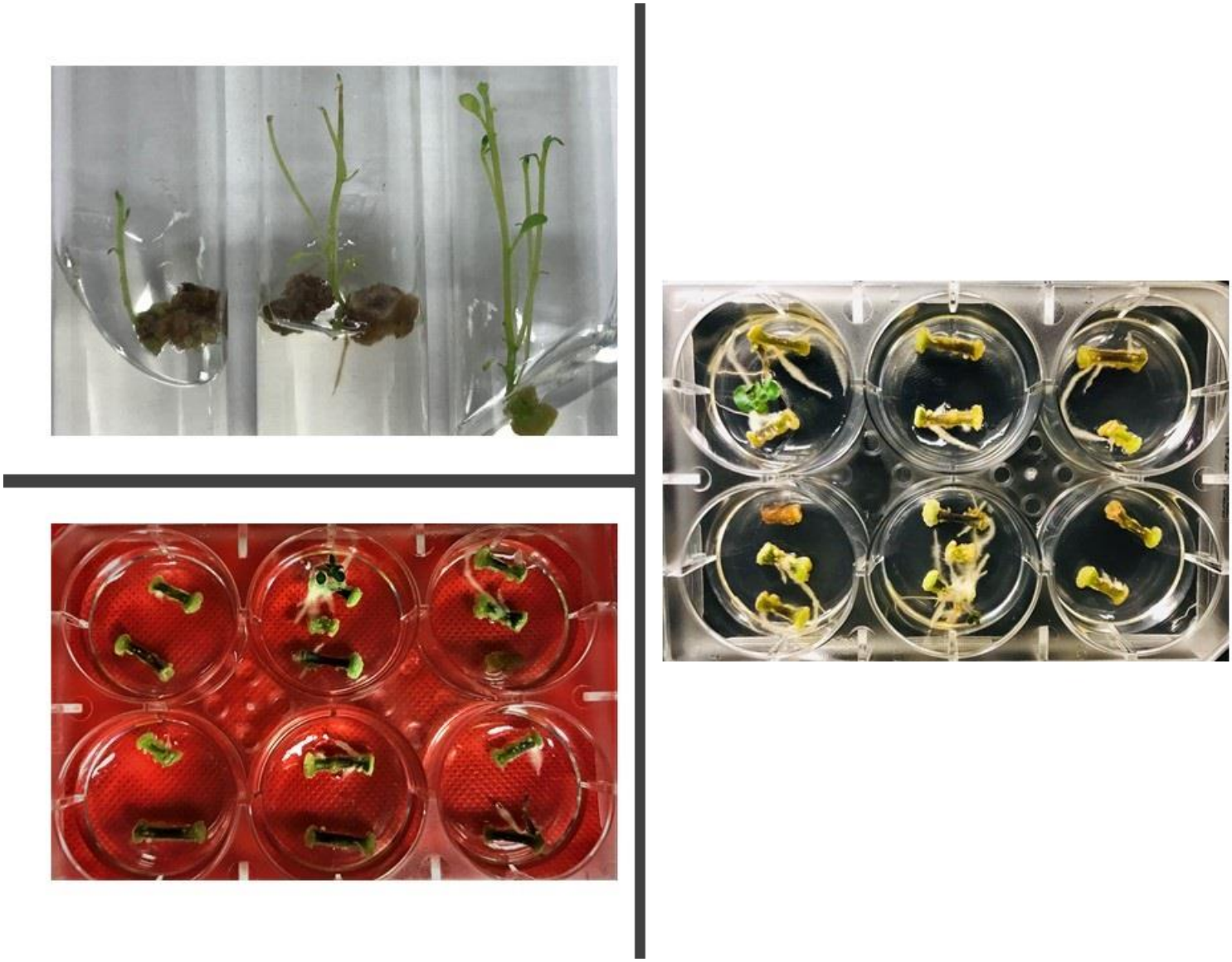
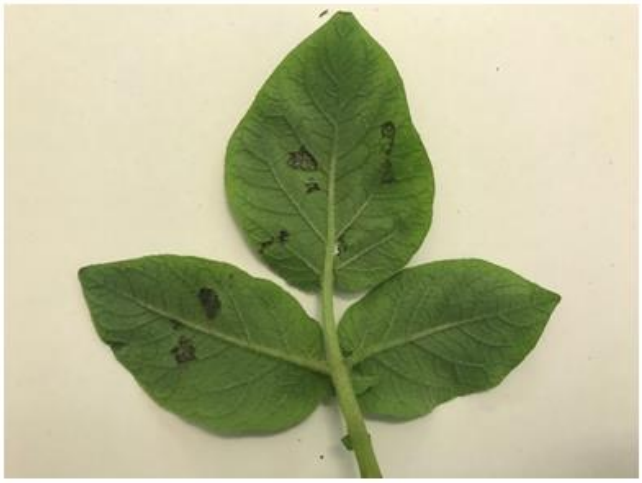


Figure 4.4: Leaf samples photographed at 6 dpi for disease severity assessment and pathogen biomass quantification.



StCCoAOMT KI

Russet Burbank

StCCoAOMT KO

F06037

Figure 4.5: *StCCoAOMT* gene (transcript variants) knock-out in late blight resistance potato genotype (F06037). **a** *StCCoAOMT* gene and transcript variants (PGSC0003DMT400006134: *StCCoAOMT* $^{p.\ Glu80*}$ and PGSC0003DMT400006135; *StCCoAOMT* $^{p.\ Glu42*}$) on chromosome 4. Two transcript variants with a protein-coding region (black boxes) and stop-gain mutation (in Russet Burbank) showed with a red bar. Two sgRNAs are designed to target both transcript variants. sgRNA target sites are underlined with the blue bar, PAM sites and target sites are marked with green and yellow colour. **b** Chromatogram from sequencing data on F06037 wild type and Homozygous *StCCoAOMT*^{KO} F06037 plants (progenies of T0 plants or first clonal generation (CG1). Red boxes adjacent to the PAM sites on a chromatogram is representing deletions at target sites in *StCCoAOMT*^{KO} F06037 plants.

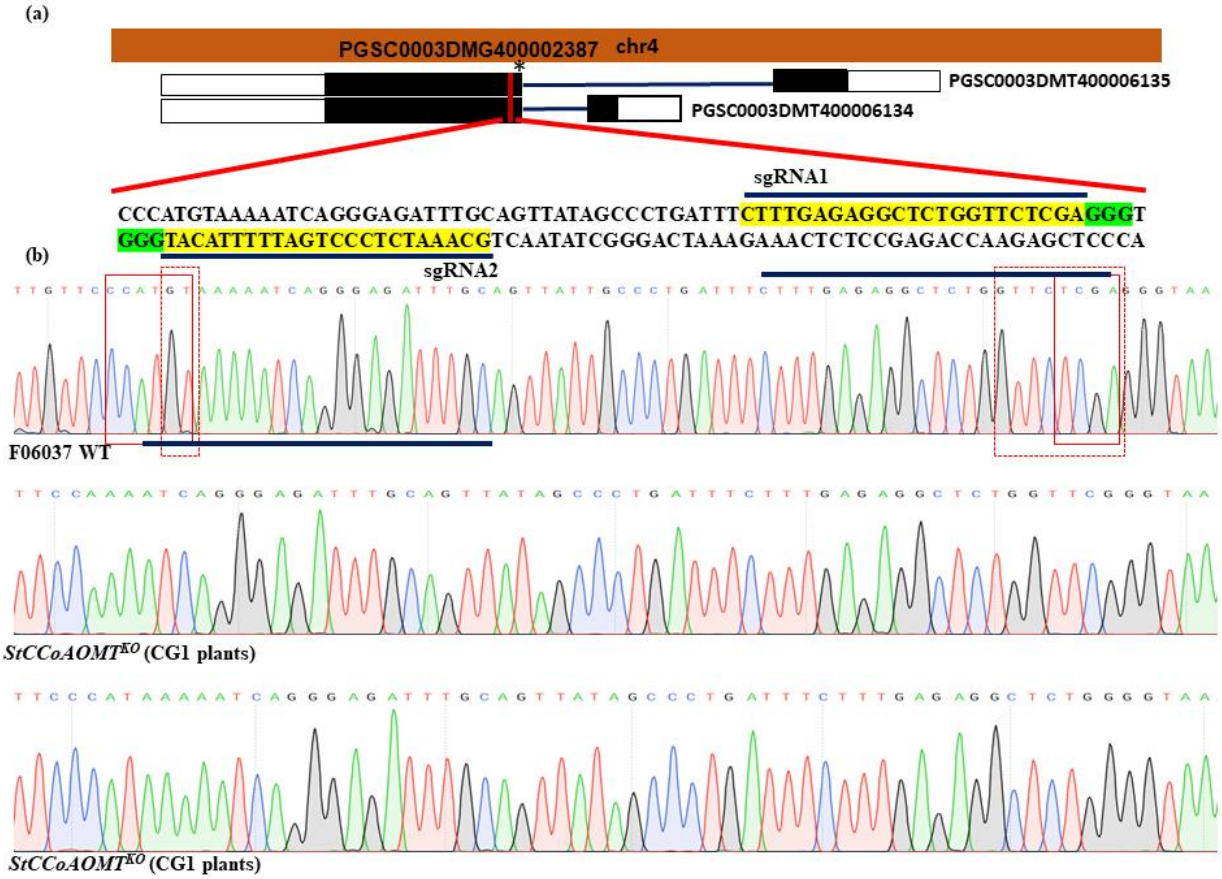
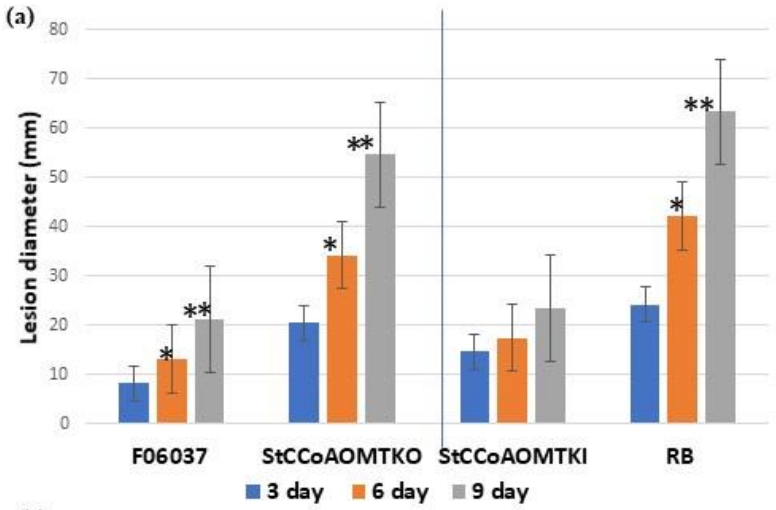
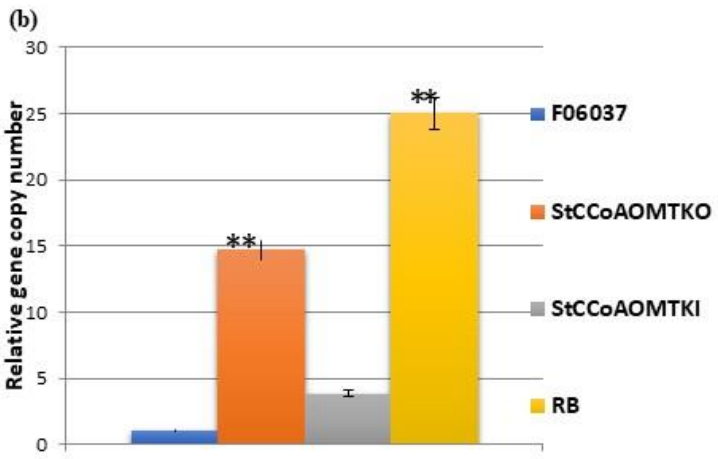
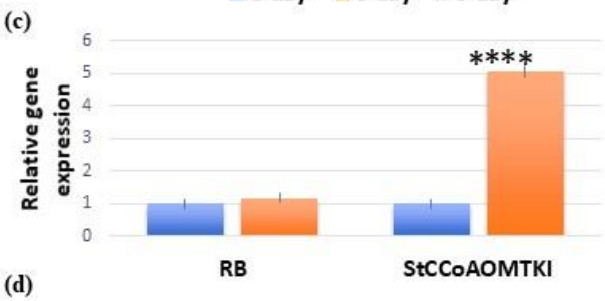
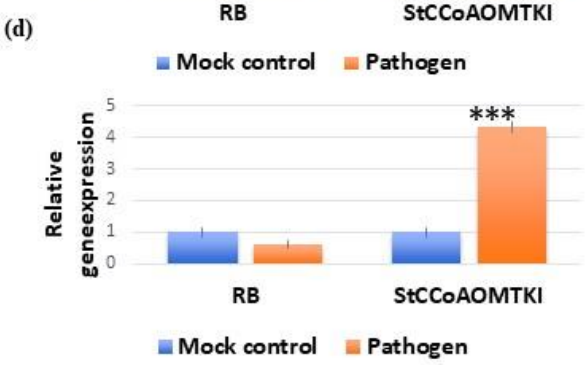


Figure 4.6: Validation of *StCCoAOMT*^{KO} and *StCCoAOMT*^{KI} plants. **a** late blight disease progress curve based on lesion diameter (mm) at 3, 6, and 9 days of post-inoculation. **b** pathogen biomass quantification at 6 days of post-inoculation (6 dpi). Pathogen biomass was quantified as relative *P. infestans*-specific (O-8) gene expression. Significant differences between (F06037 and *StCCoAOMT*^{KO} plants) and (Russet Burbank and *StCCoAOMT*^{KI} plants) using Student's *t*-test are indicated: *, $P < 0.05$; **, $P < 0.01$. **c** and **d** qRT-PCR expression studies of transcript variants (PGSC0003DMT400006134 and PGSC0003DMT400006135) in Russet Burbank and *StCCoAOMT*^{KI} RB plants following *Phytophthora infestans* inoculation, at 2 days post-inoculation (2 dpi). Comparative expression of transcripts in mock (no pathogen; control) and pathogen inoculated leaf samples of Russet Burbank and *StCCoAOMT*^{KI} RB normalized to the expression data of b-tubulin reference gene. Significant differences in expression levels of knock-in plants compared with unedited Russet Burbank genotype using Student's t-test are indicated: *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. **e** detached leaf assay showing significant late blight severity difference between wild type (RB) and *StCCoAOMT*^{KI}. Images were taken on 6 days of post-inoculation (6 dpi).


(a)
Lesion diameter (mm)
F06037
StCCoAOMTKO
StCCoAOMTKI
RB
3 day
6 day
9 day


(b)
Relative gene copy number
F06037
StCCoAOMTKO
StCCoAOMTKI
RB


(c)
Relative gene expression
RB
StCCoAOMTKI
Mock control
Pathogen


(d)
Relative geneexpression
RB
StCCoAOMTKI
Mock control
Pathogen

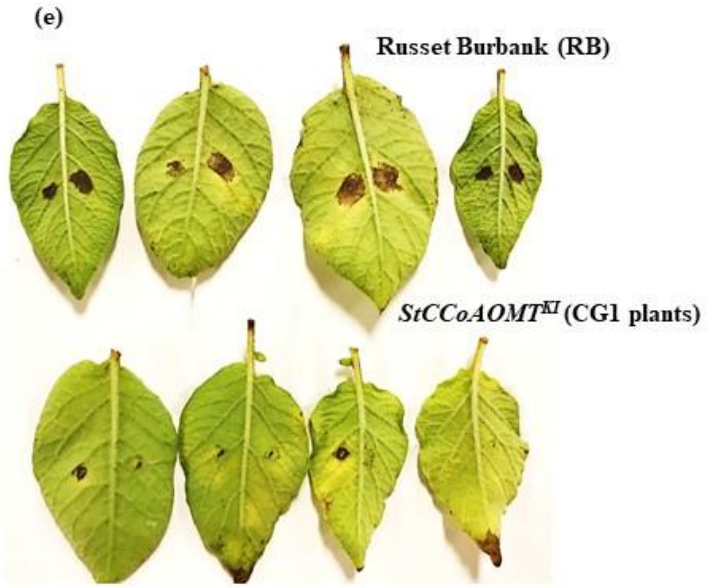

(e)
Russet Burbank (RB)
StCCoAOMTKI (CG1 plants)

Figure 4.7: Targeted gene editing on *StCCoAOMT* gene in F06037 and Russet Burbank genotypes. Protoplast transformation and efficiency of CRISPR-Cas9: Mutation analysis by cloning, sequencing of sgRNAs target sites and ICE v2 CRISPR Analysis Tool (Synthego). Different mutant alleles (allelic composition) obtained in four sets of protoplast transformation experiments are given **a** to **d**.

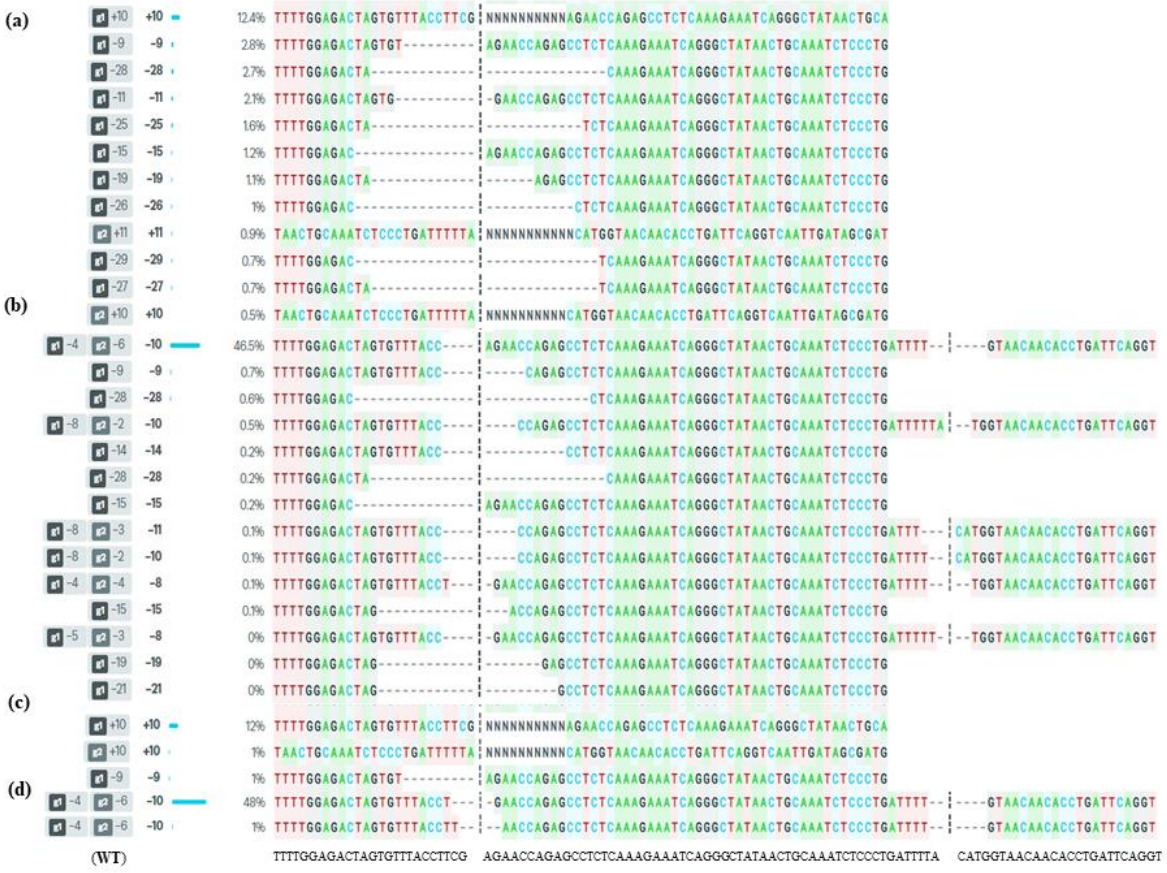


Figure 4.8: a to **d**. Examples of mutated alleles identified in CG1 plants (within target sites of both sgRNAs). Genotyping of alleles by cloning and Sanger sequencing in CG1 plants and analyzed using ICE v2 CRISPR Analysis Tool (Synthego). WT, wild type. InDel mutations (Insertion/deletion) are shown in red along with the cleavage site.

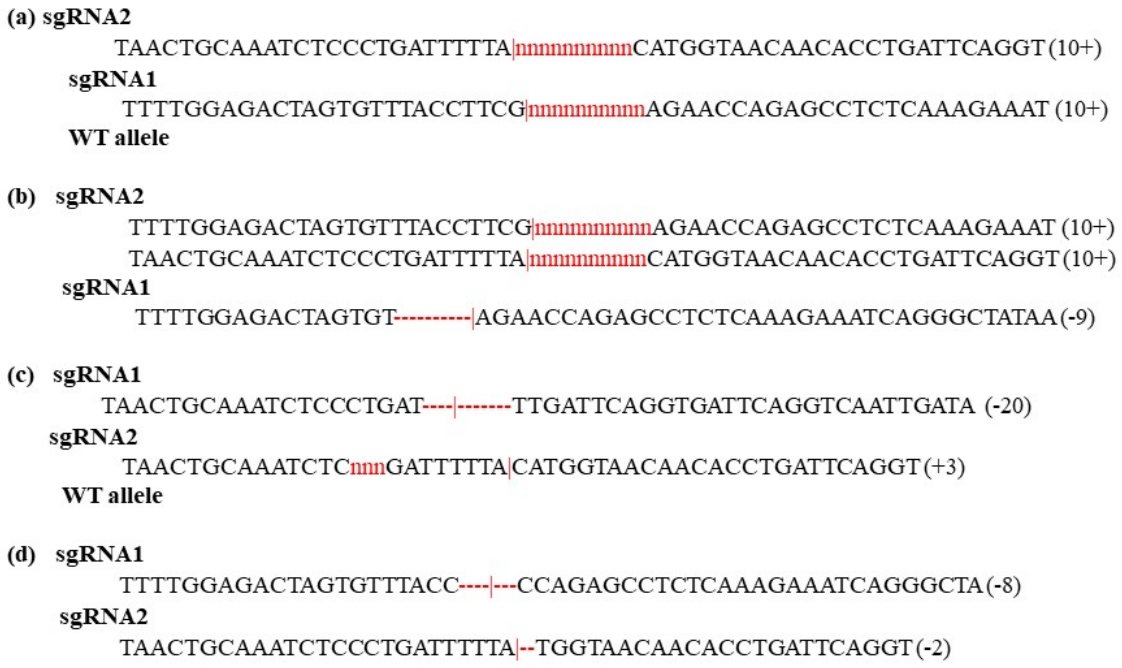


Figure 4.9: Examples of ICE CRISPR analysis tool results for *StCCoAOMT* edited Clonal generation 1(CG1) plants.

Figure 4.10: Homology Directed Repair template (HDR) design and molecular characterization of edited CG1 plants. **a** single guide RNAs (sgRNAs) design, point non-sense mutation site (* C to A), location of the primers to characterize the edited Russet Burbank plants; Schematic representation of HDR template design to carry out *StCCoAOMT* gene knock-in or mutation correction: left and right homology arms (LHA and RHA) with desired changes including base pair change (G to A) to erase *XhoI* site to facilitate the identification of HDR template integration in Russet Burbank edited (*StCCoAOMT*^{KI}) plants. MP1 and MP2 are mutated adjacent protospacer motif (PAM). **b** PCR restriction enzyme assay (PCR-RE) using PCR primers (Table S1) and *XhoI* restriction enzyme: example of screening plants for the successful integration of HDR template (HDR allele) using PCR amplification and restriction digestion with *Xho I* (C4 and C6 with HDR allele and others with unedited wild type allele).

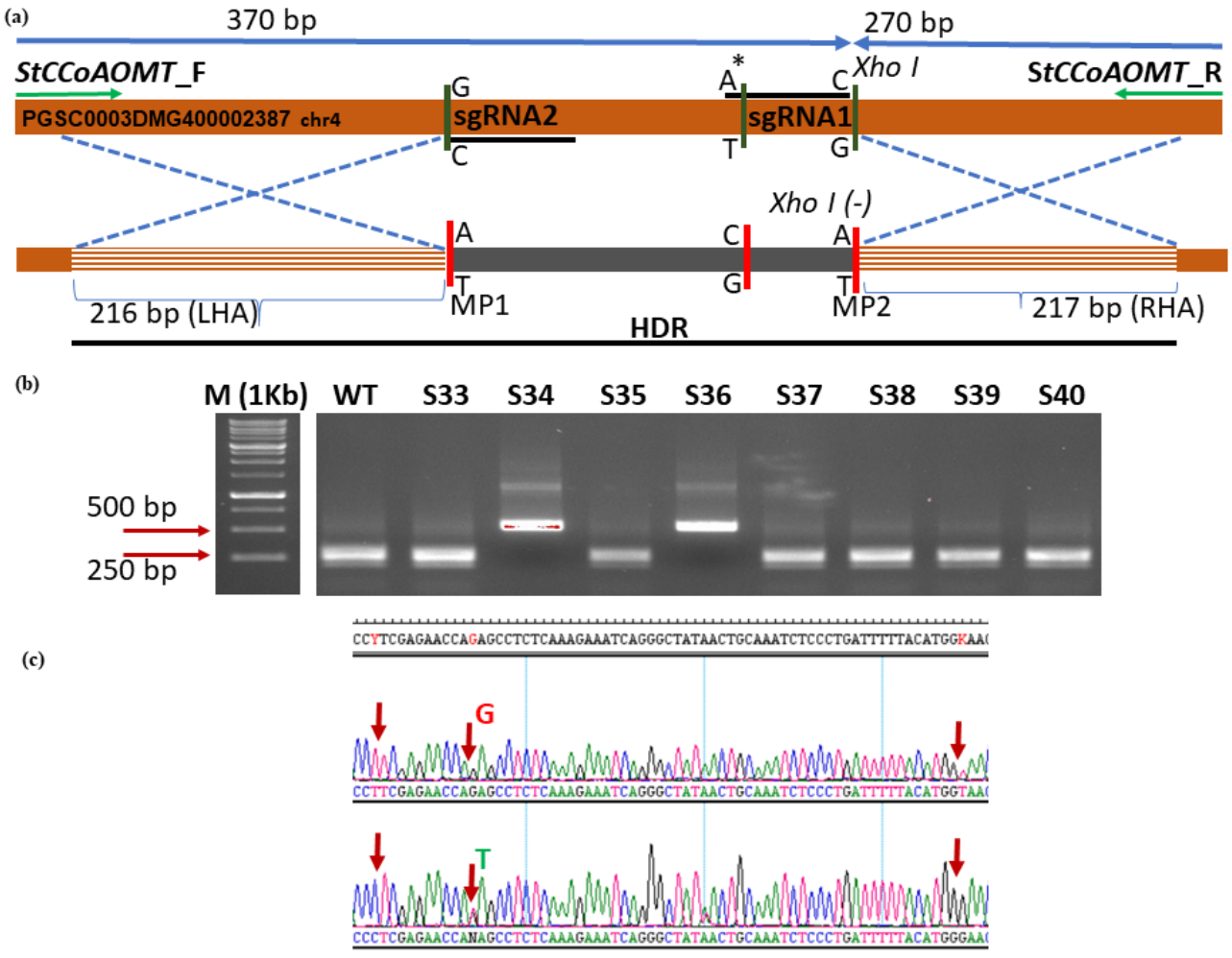
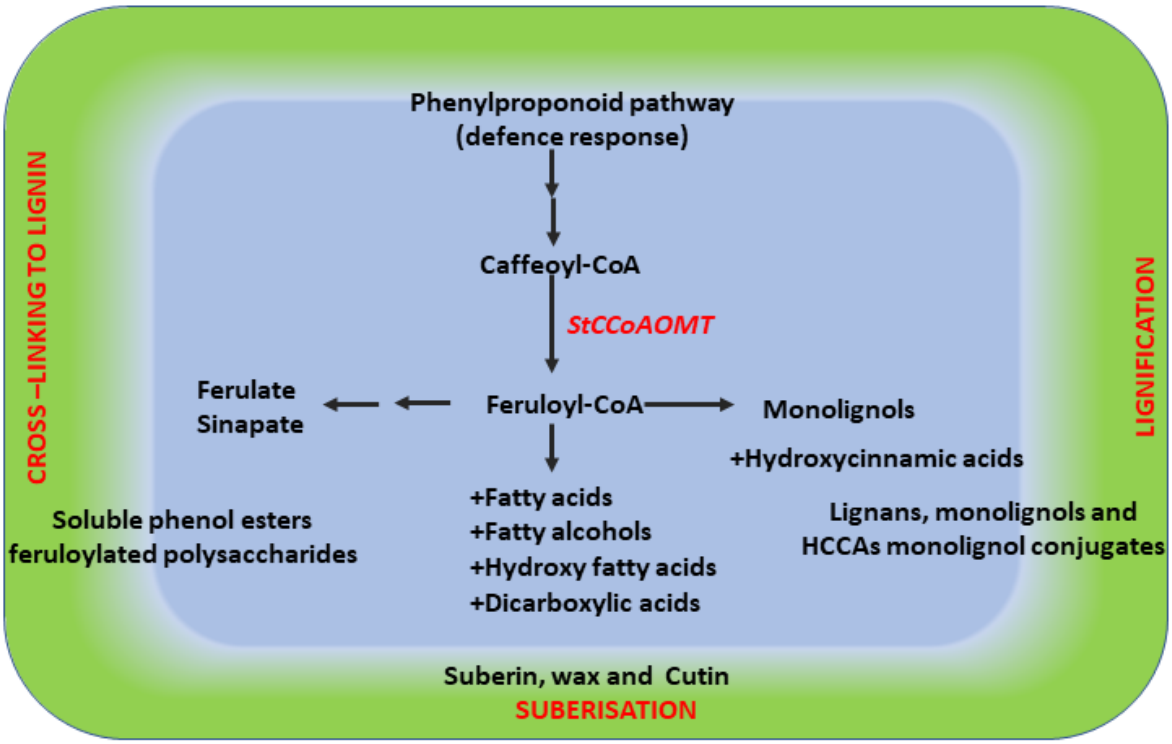


Figure 4.11: Model: Metabolic pathway showing possible secondary cell wall thickening due to the increase in the pool of feruloyl-CoA in potato in response to *P. infestans*.



CHAPTER V: GENERAL DISCUSSIONS, SUMMARY AND SUGGESTIONS FOR FUTURE RESESRCH

6.1 General discussion and summary

6.1.1 Importance of enhancing multiple disease resistance in potato

Potato (*Solanum tuberosum* L.) is one of the important staple food crops in the world with an annual production of over 300 million tons globally (FAO 2017), contributing to food and nutritional security. Potato cultivation is increasing in land area in the developing world as a cash crop for millions of farmers (Devaux et al. 2014). However, potato production worldwide is certainly hampered by various biotic and abiotic stresses (Evers et al., 2010; Hijmans, 2003). Potato production is affected by many bacterial, fungal diseases and insect pests across the potato growing countries (Were et al., 2013; Arora et al., 2014; Balendres et al., 2016). Diseases in cultivating potato varieties worsen the yield and tuber quality. Regardless of host and pathogen, resistance is mainly due to the metabolites (Kushalappa et al. 2016a). Production of a specific set of pathogen-induced metabolites in a particular plant species can able to resist multiple diseases (Piasecka et al., 2015; Wiesner-Hanks and Nelson, 2016). Apart from the conventional potato disease management strategies including fungicide application, there are limited tools available for crop disease management including potato (Rinaldo and Ayliffe 2015). The development of cultivars with improved disease resistance based on conventional breeding to manage multiple diseases is very challenging. Though the mechanisms to perceive the pathogens may vary a lot among different microbial pathogens the eventual mode of suppressing the pathogen to contain it to the initial infection area appears to be mainly due to resistance-related metabolites.

6.1.2 Discovering candidate resistance genes using RNA-sequence data

There is a need for a method that enables the identification of genes involved in triggering the biosynthesis of resistance-related metabolites that suppress the pathogen due to their antimicrobial activities. Genetic, especially the allelic variation, is the main reason for the biosynthesis of variable amounts of RR metabolites (Huang et al., 2012; Kerwin et al., 2015). Deleterious mutation within the biosynthetic and regulatory genes might greatly affect the pathogen-induced

metabolites in elite potato cultivars. Based on the advancements in Next Generation Sequencing (NGS), especially with RNA-sequencing (RNA-seq) or transcriptomic studies, it is possible to discover transcript variants of biosynthetic and regulatory genes that are induced upon pathogen attack. several tools can be used to identify the deleterious mutations within the coding region of genes (Lopez-Maestre et al. 2016; Mosquera et al. 2016). The genetic variation among the genes and the study can later be linked to the metabolite biosynthetic pathway genes in a specific potato genotype. Accordingly, several bioinformatics tools were used in chapter III to discover the candidate genes to unravel the reason for less accumulation of specific sets of metabolites in Russet Burbank upon pathogen inoculation. This discovery was started with the identity SNPs and InDels within biosynthetic pathways genes of three genotypes (susceptible to late blight: Russet Burbank; resistance: F06037 and F06025). The RNA-seq from pathogen inoculated samples were analysed to identify genetic variation among the pathogen induced genes and respective transcripts. Genes and transcripts with the deleterious mutations were identified in Russet Burbank and compared with late blight resistance genotype. The identified and selected SNPs/InDels of several genes were confirmed by Sanger sequencing and expression studies. Finally, several biosynthetic genes with deleterious mutation were identified in Russet Burbank genotype by linking their relationship with the metabolites, which were used to reinforce the secondary cell walls around the infection site, to reduce further advancement of the pathogen.

6.1.3 Exploiting genome editing tools in potato to enhance disease resistance

Further, CRISPR-Cas9 based gene editing and base editing have been opened new avenues to facilitate crop improvement including enhancing disease resistance in potato (Wang et al. 2015; Andersson et al. 2018; Veillet et al. 2019a). CRISPR-Cas9 based gene editing and base editing are promising novel tools in plant breeding with the potential to modify mutated bases or replace non-functional alleles with functional alleles of genes.

In chapter IV, one of the genes (transcripts) with deleterious mutation (*StCCoAOMT*) identified in chapter III was selected. The gene encoding caffeoyl-CoA O-methyltransferase (*StCCoAOMT*), which methylates caffeoyl-CoA to feruloyl-CoA and 5- hydroxyferuloyl-CoA to sinapoyl-CoA was selected as the candidate for gene editing. The gene plays a role in the synthesis of monolignol,

feruloylated polysaccharides, the cell wall-bound ferulic acid polymers. It is involved in the reinforcement of the cell wall in response to pathogens. The single nucleotide polymorphism (SNP) was identified in Russet Burbank genotype (chapter III). The nonsense mutation (C to A) within the two transcripts of *StCCoAOMT* gene changes the amino acid to a stop codon, thus truncating the StCCoAOMT protein. Firstly, in chapter IV, the late blight resistance function of two transcripts of the gene was validated using CRISPR-Cas9 gene knock-out in a resistant genotype, F06037. Secondly, having confirmed the function of this gene and the transcripts, the mutation in this gene was corrected based on CRISPR-Cas9. The study successfully utilized Geminivirus-replicon based CRISPR/Cas9 System to carry out gene targeting or mutation correction of the *StCCoAOMT* gene in the Russet Burbank genotype. The internodes explants were transformed with *Agrobacterium* plasmids carrying CRISPR reagents. The regenerating calli were screened for corrected copy of the gene. The gene replacement or mutation correction was confirmed by restriction digestion assays followed by sanger sequencing. T0 and CG1 (clonal generation 1) edited plants showed decrease in disease severity and pathogen biomass. The metabolic profiling of the edited plants showed the significant accumulation of feruloylated metabolites. Which are known to have roles in secondary cell wall reinforcement to contain the pathogen (Yogendra et al. 2015a; Kumar et al. 2016a).

6.2 Suggested future studies

- Candidate genes identified in the study can be sequenced in other potato genotypes. If the genes and transcripts are mutated, they can be associated with resistance-related metabolite production and disease resistance in respective genotypes.
- The mutated alleles of the genes identified in the study can be replaced with functional alleles of the genes using the CRISPR-Cas9 genome editing tool. Enhanced disease resistance can be proved, if this genotype is associated with the rest of the gene hierarchy involved in the metabolic pathway to biosynthesize the RR metabolites.
- These candidates with deleterious mutations can be studied in different potato genotypes for multiple disease resistance
- Genes with mutated alleles (like *StCCoAOMT*) can be corrected using other genome editing tools including base editing to achieve transgene-free gene editing, thus one of the major regulatory requirements of Canada can be met.

RFERENCES

Agrios, G. N. (1988). Plant pathology. San Diego etc.: Acad. press.

Al-Haggar M, Madej-Pilarczyk A, Kozlowski L, et al (2012) A novel homozygous p.Arg527Leu LMNA mutation in two unrelated Egyptian families causes overlapping mandibuloacral dysplasia and progeria syndrome. European Journal of Human Genetics 20:1134–1140. https://doi.org/10.1038/ejhg.2012.77

Almagro L, Gómez Ros L V, Belchi-Navarro S, et al (2008) Class III peroxidases in plant defence reactions. Journal of Experimental Botany 60:377–390. https://doi.org/10.1093/jxb/ern277

Altshuler D, Pollara VJ, Cowles CR, et al (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature 407:513

Andersen EJ, Ali S, Byamukama E, et al (2018) Disease Resistance Mechanisms in Plants. Genes 9:339. https://doi.org/10.3390/genes9070339

Andersson M, Turesson H, Nicolia A, et al (2017) Efficient targeted multiallelic mutagenesis in tetraploid potato (Solanum tuberosum) by transient CRISPR-Cas9 expression in protoplasts. Plant Cell Reports 36:117–128. https://doi.org/10.1007/s00299-016-2062-3

Andersson M, Turesson H, Olsson N, et al (2018) Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. Physiologia Plantarum 164:378–384. https://doi.org/10.1111/ppl.12731

Arencibia AD, Carmona ER, Tellez P, et al (1998) An efficient protocol for sugarcane (Saccharum spp. L.) transformation mediated by Agrobacterium tumefaciens. Transgenic Research 7:213–222. https://doi.org/10.1023/A:1008845114531

Arora R, Sharma S, Singh B (2014) Late blight disease of potato and its management. Potato Journal. 41(1).

Asai S, Ohta K, Yoshioka H (2008) MAPK Signaling Regulates Nitric Oxide and NADPH Oxidase-Dependent Oxidative Bursts in Nicotiana benthamiana. The Plant Cell 20:1390–1406. https://doi.org/10.1105/tpc.107.055855

Axelstad M, Boberg J, Nellemann C, et al (2011) Exposure to the Widely Used Fungicide

Mancozeb Causes Thyroid Hormone Disruption in Rat Dams but No Behavioral Effects in the Offspring. Toxicological Sciences 120:439–446

Aylor DE, Schmale DG, Shields EJ, et al (2011) Tracking the potato late blight pathogen in the atmosphere using unmanned aerial vehicles and Lagrangian modeling. Agricultural and Forest Meteorology 151:251–260. https://doi.org/10.1016/j.agrformet.2010.10.013

Balendres MA, Tegg RS, Wilson CR (2016) Key events in pathogenesis of spongospora diseases in potato: a review. Australasian Plant Pathology 45:229–240. https://doi.org/10.1007/s13313-016-0398-3

Ballvora A, Ercolano MR, Weiss J, et al (2002) The R1 gene for potato resistance to late blight (Phytophthora infestans) belongs to the leucine zipper/NBS/LRR class of plant resistance genes. The Plant Journal 30:361–371. https://doi.org/10.1046/j.1365-313X.2001.01292.x

Baltes NJ, Gil-Humanes J, Cermak T, et al (2014) DNA Replicons for Plant Genome Engineering. The Plant Cell 26:151–163. https://doi.org/10.1105/tpc.113.119792

Bassard J-E, Richert L, Geerinck J, et al (2012) Protein–Protein and Protein–Membrane Associations in the Lignin Pathway. The Plant Cell 24:4465 LP – 4482. https://doi.org/10.1105/tpc.112.102566

Belmaaza A, Chartrand P (1994) One-sided invasion events in homologous recombination at double-strand breaks. Mutation Research-DNA Repair 314:199–208

Benton HP, Wong DM, Trauger SA, Siuzdak G (2008) XCMS ^{2}: Processing Tandem Mass Spectrometry Data for Metabolite Identification and Structural Characterization. Analytical Chemistry 80:6382–6389. https://doi.org/10.1021/ac800795f

Biała W, Jasiński M (2018) The Phenylpropanoid Case – It Is Transport That Matters . Frontiers in Plant Science 9:1610

Boch J, Scholze H, Schornack S, et al (2009) Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science 326:1509 LP – 1512

Bollina V, Kumarswamy GK, Kushalappa AC, et al (2010) Mass spectrometry-based metabolomics application to identify quantitative resistance-related metabolites in barley

against Fusarium head blight. Molecular Plant Pathology 11: https://doi.org/10.1111/j.1364-3703.2010.00643.x

Bollina V, Kushalappa AC, Choo TM, et al (2011) Identification of metabolites related to mechanisms of resistance in barley against Fusarium graminearum, based on mass spectrometry. Plant Molecular Biology 77:355–370. https://doi.org/10.1007/s11103-011-9815-8

Burlingame B, Mouillé B, Charrondière R (2009) Nutrients, bioactive non-nutrients and anti-nutrients in potatoes. Journal of Food Composition and Analysis 22:494–502. https://doi.org/10.1016/j.jfca.2009.09.001

Burra DD, Berkowitz O, Hedley PE, et al (2014) Phosphite-induced changes of the transcriptome and secretome in Solanum tuberosum leading to resistance against Phytophthora infestans. BMC Plant Biology 14:254. https://doi.org/10.1186/s12870-014-0254-y

Burris KP, Dlugosz EM, Collins AG, et al (2016) Development of a rapid, low-cost protoplast transfection system for switchgrass (Panicum virgatum L.). Plant Cell Reports 35:693–704. https://doi.org/10.1007/s00299-015-1913-7

Butler NM, Baltes NJ, Voytas DF, Douches DS (2016) Geminivirus-Mediated Genome Editing in Potato (Solanum tuberosum L.) Using Sequence-Specific Nucleases. Frontiers in Plant Science 7:1045. https://doi.org/10.3389/fpls.2016.01045

Čermák T, Baltes NJ, Čegan R, et al (2015) High-frequency, precise modification of the tomato genome. Genome Biology 16:232. https://doi.org/10.1186/s13059-015-0796-9

Chen J, Ullah C, Reichelt M, et al (2019a) Sclerotinia sclerotiorum Circumvents Flavonoid Defenses by Catabolizing Flavonol Glycosides and Aglycones. Plant Physiology 180:1975 LP – 1987. https://doi.org/10.1104/pp.19.00461

Chen X, Wang H, Li X, et al (2019b) Molecular cloning and functional analysis of 4-Coumarate:CoA ligase 4(4CL-like 1)from Fraxinus mandshurica and its role in abiotic stress tolerance and cell wall synthesis. BMC plant biology 19:231. https://doi.org/10.1186/s12870-019-1812-0

Cheng Q, Li N, Dong L, et al (2015) Overexpression of Soybean Isoflavone Reductase (GmIFR)

Enhances Resistance to Phytophthora sojae in Soybean. Frontiers in plant science 6:1024. https://doi.org/10.3389/fpls.2015.01024

Cho K, Cho K-S, Sohn H-B, et al (2016) Network analysis of the metabolome and transcriptome reveals novel regulation of potato pigmentation. Journal of Experimental Botany 67:1519–1533. https://doi.org/10.1093/jxb/erv549

Cho SW, Kim S, Kim Y, et al (2014) Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research 24:132–141. https://doi.org/10.1101/gr.162339.113

CIP (2015). International potato center, Agricultural research for development.

Christian M, Cermak T, Doyle EL, et al (2010) Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. Genetics 186:757 LP – 761

Cingolani P, Platts A, Wang LL, et al (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80–92. https://doi.org/10.4161/fly.19695

Clasen B, Stoddard T, Luo S, et al (2015) Improving cold storage and processing traits in potato through targeted gene knockout. Plant biotechnology journal 14: https://doi.org/10.1111/pbi.12370

Clevenger J, Chavarro C, Pearl SA, et al (2015) Single Nucleotide Polymorphism Identification in Polyploids: A Review, Example, and Recommendations. Molecular Plant 8:831–846. https://doi.org/https://doi.org/10.1016/j.molp.2015.02.002

Colin A. Smith, Elizabeth J. Want, Grace O'Maille, et al (2006) XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. https://doi.org/10.1021/AC051437Y

Collonnier C, Guyon-Debast A, Maclot F, et al (2017) Towards mastering CRISPR-induced gene knock-in in plants: Survey of key features and focus on the model Physcomitrella patens. Methods 121–122:103–117. https://doi.org/https://doi.org/10.1016/j.ymeth.2017.04.024

Cong L, Ran FA, Cox D, et al (2013) Multiplex genome engineering using CRISPR/Cas systems. Science (New York, NY) 339:819–23. https://doi.org/10.1126/science.1231143

Cooke LR, Schepers HTAM, Hermansen A, et al (2011) Epidemiology and Integrated Control of Potato Late Blight in Europe. Potato Research 54:183–222. https://doi.org/10.1007/s11540-011-9187-0

Cordovado SK, Hendrix M, Greene CN, et al (2012) CFTR mutation analysis and haplotype associations in CF patients(). Molecular genetics and metabolism 105:249–254. https://doi.org/10.1016/j.ymgme.2011.10.013

Cromme N, Prakash AB, Lutaladio N, Ezeta F (2010) Strengthening potato value chains: technical and policy options for developing countries. Food and Agriculture Organization of the United Nations (FAO)

Dahan-Meir T, Filler-Hayut S, Melamed-Bessudo C, et al (2018) Efficient in planta gene targeting in tomato using geminiviral replicons and the CRISPR/Cas9 system. The Plant Journal 95:5–16. https://doi.org/10.1111/tpj.13932

De Lacy Costello BPJ, Evans P, Ewen RJ, et al (2001) Gas chromatography-mass spectrometry analyses of volatile organic compounds from potato tubers inoculated with Phytophthora infestans or Fusarium coeruleum. Plant Pathology 50:489–496. https://doi.org/10.1046/j.1365-3059.2001.00594.x

de O. Buanafina MM (2009) Feruloylation in Grasses: Current and Future Perspectives. Molecular Plant 2:861–872. https://doi.org/https://doi.org/10.1093/mp/ssp067

De Vos RCH, Moco S, Lommen A, et al (2007) Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. Nat Protoc 2:778–791. https://doi.org/10.1038/nprot.2007.95

Deng Y, Lu S (2017) Biosynthesis and Regulation of Phenylpropanoids in Plants. Critical Reviews in Plant Sciences 36:257–290. https://doi.org/10.1080/07352689.2017.1402852

Devaux A, Kromann P, Ortiz O (2014) Potatoes for sustainable global food security. Potato Research 57:185–199

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 11:e0155851. https://doi.org/10.1371/journal.pone.0155851

Dixon RA, Achnine L, Kota P, et al (2002) The phenylpropanoid pathway and plant defence—a genomics perspective. Molecular Plant Pathology 3:371–390. https://doi.org/10.1046/j.1364-3703.2002.00131.x

Dixon RA, Paiva NL (1995) Stress-Induced Phenylpropanoid Metabolism. The Plant cell 7:1085–1097. https://doi.org/10.1105/tpc.7.7.1085

Domergue F, Vishwanath SJ, Joubès J, et al (2010) Three Arabidopsis fatty acyl-coenzyme A reductases, FAR1, FAR4, and FAR5, generate primary fatty alcohols associated with suberin deposition. Plant physiology 153:1539–1554. https://doi.org/10.1104/pp.110.158238

Domfeh O, Bittara FG, Gudmestad NC (2014) Sensitivity of Potato Cultivars to Potato mop-top virus-Induced Tuber Necrosis. Plant Disease 99:788–796. https://doi.org/10.1094/PDIS-07-14-0705-RE

Dong OX, Yu S, Jain R, et al (2020) Marker-free carotenoid-enriched rice generated through targeted gene insertion using CRISPR-Cas9. Nature Communications 11:1178. https://doi.org/10.1038/s41467-020-14981-y

Drabovich AP, Krylov SN (2006) Identification of Base Pairs in Single-Nucleotide Polymorphisms by MutS Protein-Mediated Capillary Electrophoresis. Analytical Chemistry 78:2035–2038. https://doi.org/10.1021/ac0520386

Draffehn A, Li L, Krezdorn N, et al (2013) Comparative transcript profiling by SuperSAGE identifies novel candidate genes for controlling potato quantitative resistance to late blight not compromised by late maturity. Frontiers in Plant Science 4:423

Duan H, Richael C, Rommens CM (2012) Overexpression of the wild potato eIF4E-1 variant Eva1 elicits Potato virus Y resistance in plants silenced for native eIF4E-1. Transgenic Research 21:929–938. https://doi.org/10.1007/s11248-011-9576-9

Eckardt NA (2002) Plant Disease Susceptibility Genes? The Plant Cell 14:1983–1986. https://doi.org/10.1105/tpc.140910

Enciso-Rodriguez F, Douches D, Lopez-Cruz M, et al (2018) Genomic Selection for Late Blight and Common Scab Resistance in Tetraploid Potato (Solanum tuberosum). G3 (Bethesda, Md) 8:2471–2481. https://doi.org/10.1534/g3.118.200273

Enciso-Rodriguez F, Manrique-Carpintero NC, Nadakuduti SS, et al (2019) Overcoming Self-Incompatibility in Diploid Potato Using CRISPR-Cas9. Frontiers in Plant Science 10:376

Endo A, Masafumi M, Kaya H, Toki S (2016) Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from Francisella novicida. 6:38169

Evers D, Lefèvre I, Legay S, et al (2010) Identification of drought-responsive compounds in potato through a combined transcriptomic and targeted metabolite approach. Journal of Experimental Botany 61:2327–2343. https://doi.org/10.1093/jxb/erq060

FAO (2009). Sustainable potato production GUIDELINES FOR DEVELOPING COUNTRIES

FAO (2010). Strengthening Potato Value Chains: Technical and policy options for developing countries. Food and Agriculture Organization of the United Nations and the Common Fund for Commodities. Rome, Italy.

FAO U (2014) FAOstat. Retrieved Feb 2014: FAO, I. (2014). WFP. 2014. The state of food insecurity in the world.

FAOSTAT FAO (2015) Faostat FAO, I. (2015). WFP. 2015. The state of food insecurity in the world. Gebhardt, C. (2013). Bridging the gap between genome analysis and precision breeding in potato. Trends Genet 29, 248–256.

FAO E (2017) Food and Agriculture Organization of the United Nations: Rome

Fernie AR, Schauer N (2009) Metabolomics-assisted breeding: a viable option for crop improvement? Trends in genetics 25:39–48

Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. Plant Molecular Biology 48:155–171. https://doi.org/10.1023/A:1013713905833

Fister AS, Landherr L, Maximova SN, Guiltinan MJ (2018) Transient Expression of CRISPR/Cas9 Machinery Targeting TcNPR3 Enhances Defense Response in Theobroma cacao. Frontiers in Plant Science 9:268

Flor HH (1971) Current status of the Gene-For-Gene concept. Annual Review of Phytopathology 9:275–296. https://doi.org/10.1146/annurev.py.09.090171.001423

Foolad MR, Merk HL, Ashrafi H (2008) Genetics, genomics and breeding of late blight and early

blight resistance in tomato. Critical Reviews in Plant Sciences 27:75–107

Forbes G, J. Grünwald N, Mizubuti E, Andrade-Piedra J (2019) Potato Late Blight in Developing Countries. In: Peters, R.(ed.) Current Concepts in Potato Disease Management. Kerala, India: Research Signpost (In press).

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

Fry WE, Goodwin SB (1997) Reemergence of potato and tomato late bligth in United States. Plant Desease 69–72. https://doi.org/10.1094/PDIS.1997.81.12.1349

Fry WE, McGrath MT, Seaman a., et al (2013) The 2009 late blight pandemic in Eastern USA – causes and results. Plant Disease 97:296–306. https://doi.org/10.1094/PDIS-08-12-0791-FE

Gálvez JH, Tai HH, Lagüe M, et al (2016) The nitrogen responsive transcriptome in potato (Solanum tuberosum L.) reveals significant gene regulatory motifs. Scientific reports 6:26090

Gao J, Wang G, Ma S, et al (2015) CRISPR/Cas9-mediated targeted mutagenesis in Nicotiana tabacum. Plant Molecular Biology 87:99–110. https://doi.org/10.1007/s11103-014-0263-0

Gao L, Bradeen JM (2016) Contrasting Potato Foliage and Tuber Defense Mechanisms against the Late Blight Pathogen Phytophthora infestans. PLOS ONE 11:e0159969

Garber M, Grabherr MG, Guttman M, Trapnell C (2011) Computational methods for transcriptome annotation and quantification using RNA-seq. Nat Meth 8:469–477

Gil-Humanes J, Wang Y, Liang Z, et al (2017) High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. The Plant Journal 89:1251–1262. https://doi.org/10.1111/tpj.13446

Gong L, Zhang H, Gan X, et al (2015) Transcriptome Profiling of the Potato (Solanum tuberosum L.) Plant under Drought Stress and Water-Stimulus Conditions. PLOS ONE 10:e0128041

Gou J-Y, Yu X-H, Liu C-J (2009) A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 106:18855–18860. https://doi.org/10.1073/pnas.0905555106

Goyer A, Hamlin L, Crosslin JM, et al (2015) RNA-Seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. BMC Genomics 16:472. https://doi.org/10.1186/s12864-015-1666-2

Gratz SJ, Rubinstein CD, Harrison MM, et al (2015) CRISPR-Cas9 genome editing in Drosophila. Current Protocols in Molecular Biology 2015:31.2.1-31.2.20. https://doi.org/10.1002/0471142727.mb3102s111

Griffin TJ, Smith LM (2000) Genetic Identification by Mass Spectrometric Analysis of Single-Nucleotide Polymorphisms: Ternary Encoding of Genotypes. Analytical Chemistry 72:3298–3302. https://doi.org/10.1021/ac991390e

Haas BJ, Kamoun S, Zody MC, et al (2009) Genome sequence and analysis of the Irish potato famine pathogen Phytophthora infestans. Nature 461:393–398. https://doi.org/10.1038/nature08358

Hahn F, Eisenhut M, Mantegazza O, Weber APM (2018) Homology-Directed Repair of a Defective Glabrous Gene in Arabidopsis With Cas9-Based Gene Targeting . Frontiers in Plant Science 9:424

Hall R, Beale M, Fiehn O, et al (2002) Plant metabolomics: the missing link in functional genomics strategies. The Plant cell 14:1437–40. https://doi.org/10.1105/TPC.140720

Hall RD (2006) Plant metabolomics: from holistic hope, to hype, to hot topic. New Phytologist 169:453–468. https://doi.org/10.1111/j.1469-8137.2005.01632.x

Hameed A, Iqbal Z, Asad S, Mansoor S (2014) Detection of Multiple Potato Viruses in the Field Suggests Synergistic Interactions among Potato Viruses in Pakistan. The Plant Pathology Journal 30:407–415. https://doi.org/10.5423/PPJ.OA.05.2014.0039

Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009a) Applied biotechnology to combat late blight in potato caused by phytophthora infestans. Potato Research 52:249–264. https://doi.org/10.1007/s11540-009-9136-3

Haverkort AJ, Struik PC, Visser RGF, Jacobsen E (2009b) Applied biotechnology to combat late blight in potato caused by phytophthora infestans. Potato Research 52:249–264. https://doi.org/10.1007/s11540-009-9136-3

Hegde N, Doddamani D, Kushalappa A (2020) Identification and functional characterization of late blight resistance polymorphic genes in Russet Burbank potato cultivar Functional Plant Biology (Accepeted)

Hijmans RJ (2003) The effect of climate change on global potato production. American Journal of Potato Research 80:271–279. https://doi.org/10.1007/BF02855363

Hirsch CD, Hamilton JP, Childs KL, et al (2014) Spud DB: A Resource for Mining Sequences, Genotypes, and Phenotypes to Accelerate Potato Breeding. The Plant Genome 7:0. https://doi.org/10.3835/plantgenome2013.12.0042

Huang Y-F, Doligez A, Fournier-Level A, et al (2012) Dissecting genetic architecture of grape proanthocyanidin composition through quantitative trait locus mapping. BMC Plant Biology 12:30. https://doi.org/10.1186/1471-2229-12-30

Iliakis G, Wang H, Perrault AR, et al (2004) Mechanisms of DNA double strand break repair and chromosome aberration formation. Cytogenetic and Genome Research 104:14–20. https://doi.org/10.1159/000077461

Irani S, Todd CD, Wei Y, Bonham-Smith PC (2019) Changes in phenylpropanoid pathway gene expression in roots and leaves of susceptible and resistant Brassica napus lines in response to Plasmodiophora brassicae inoculation. Physiological and Molecular Plant Pathology 106:196–203. https://doi.org/https://doi.org/10.1016/j.pmpp.2019.02.007

Ittisoponpisan S, Islam SA, Khanna T, et al (2019) Can Predicted Protein 3D Structures Provide Reliable Insights into whether Missense Variants Are Disease Associated? Journal of Molecular Biology 431:2197–2212. https://doi.org/https://doi.org/10.1016/j.jmb.2019.04.009

Jiang H, Wong WH (2009) Statistical inferences for isoform expression in RNA-Seq. Bioinformatics 25:. https://doi.org/10.1093/bioinformatics/btp113

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444:323–329. https://doi.org/10.1038/nature05286

Judelson HS (1997) The Genetics and Biology ofPhytophthora infestans:Modern Approaches to a Historical Challenge. Fungal Genetics and Biology 22:65–76.

https://doi.org/https://doi.org/10.1006/fgbi.1997.1006

Katajamaa M, Miettinen J, Orešič M (2006) MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. Bioinformatics 22:634–636. https://doi.org/10.1093/bioinformatics/btk039

Kerwin R, Feusier J, Corwin J, et al (2015) Natural genetic variation in Arabidopsis thaliana defense metabolism genes modulates field fitness. eLife 4:e05604. https://doi.org/10.7554/eLife.05604

Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. Nature Methods 12:357

Kim D, Pertea G, Trapnell C, et al (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biology 14:R36–R36. https://doi.org/10.1186/gb-2013-14-4-r36

Kosma DK, Molina I, Ohlrogge JB, Pollard M (2012) Identification of an Arabidopsis fatty alcohol:caffeoyl-Coenzyme A acyltransferase required for the synthesis of alkyl hydroxycinnamates in root waxes. Plant physiology 160:237–248. https://doi.org/10.1104/pp.112.201822

Kou Y, Wang S (2010) Broad-spectrum and durability: understanding of quantitative disease resistance. Current Opinion in Plant Biology 13:181–185

Kumar A, Yogendra KN, Karre S, et al (2016a) WAX INDUCER1 (HvWIN1) transcription factor regulates free fatty acid biosynthetic genes to reinforce cuticle to resist Fusarium head blight in barley spikelets. Journal of Experimental Botany 67:4127–4139. https://doi.org/10.1093/jxb/erw187

Kumar S, Stecher G, Tamura K (2016b) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 33:1870–1874. https://doi.org/10.1093/molbev/msw054

Kumaraswamy GK, Bollina V, Kushalappa AC, et al (2011) Metabolomics technology to phenotype resistance in barley against Gibberella zeae. European Journal of Plant Pathology 130:29–43. https://doi.org/10.1007/s10658-010-9729-3

Kumlay AM, Ercisli S (2015) Callus induction, shoot proliferation and root regeneration of potato (Solanum tuberosum L.) stem node and leaf explants under long-day conditions. Biotechnology & Biotechnological Equipment 29:1075–1084. https://doi.org/10.1080/13102818.2015.1077685

Kusano H, Ohnuma M, Mutsuro-Aoki H, et al (2018) Establishment of a modified CRISPR/Cas9 system with increased mutagenesis frequency using the translational enhancer dMac3 and multiple guide RNAs in potato. Scientific Reports 8:13753. https://doi.org/10.1038/s41598-018-32049-2

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

Kushalappa AC, Yogendra KN, Karre S (2016a) Plant Innate Immune Response: Qualitative and Quantitative Resistance. Critical Reviews in Plant Sciences 35:38–55. https://doi.org/10.1080/07352689.2016.1148980

Kushalappa AC, Yogendra KN, Sarkar K, et al (2016b) Gene discovery and genome editing to develop cisgenic crops with improved resistance against pathogen infection. Canadian Journal of Plant Pathology 07060661.2016.1199597. https://doi.org/10.1080/07060661.2016.1199597

L. Deahl K, Inglis DA, P. DeMuth S (1993) Testing for resistance to metalaxyl in Phytophthora infestans isolates from Northwestern Washington

Lai Z, Kane NC, Kozik A, et al (2012) Genomics of Compositae weeds: EST libraries, microarrays, and evidence of introgression. American Journal of Botany 99:209–218. https://doi.org/10.3732/ajb.1100313

Langmead B, Hansen KD, Leek JT (2010) Cloud-scale RNA-sequencing differential expression analysis with Myrna. Genome Biol 11:. https://doi.org/10.1186/gb-2010-11-8-r83

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:. https://doi.org/10.1186/gb-2009-10-3-r25

Lauvergeat V, Lacombe E, Lasserre E, et al (2001) Two cinnamoyl CoA reductase (CCR) genes

from Arabidopsis are differentially expressed during development and in response to infection with pathogenic bacteria. Phytochemistry 57:1187–1195. https://doi.org/10.1016/S0031-9422(01)00053-X

Lawrenson T, Shorinola O, Stacey N, et al (2015) Induction of targeted, heritable mutations in barley and Brassica oleracea using RNA-guided Cas9 nuclease. Genome Biology 16:258. https://doi.org/10.1186/s13059-015-0826-7

Li D, Zeng R, Li Y, et al (2016) Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F1 hybrids. Scientific Reports 6:24984

Li G, Pan T, Guo D, Li L-C (2014) Regulatory Variants and Disease: The E-Cadherin −160C/A SNP as an Example. Molecular Biology International 2014:967565. https://doi.org/10.1155/2014/967565

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27:2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A, et al (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Li J-F, Norville JE, Aach J, et al (2013a) Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. Nature Biotechnology 31:688–691. https://doi.org/10.1038/nbt.2654

Li J, Zhang X, Sun Y, et al (2018) Efficient allelic replacement in rice by gene editing: A case study of the NRT1.1B gene. Journal of Integrative Plant Biology 60:536–540. https://doi.org/10.1111/jipb.12650

Li X, Chen W, Zhao Y, et al (2013b) Downregulation of caffeoyl-CoA O-methyltransferase (CCoAOMT) by RNA interference leads to reduced lignin production in maize straw. Genetics and molecular biology 36:540–546. https://doi.org/10.1590/S1415-47572013005000039

Li Y, Kim JI, Pysh L, Chapple C (2015) Four Isoforms of Arabidopsis 4-Coumarate:CoA Ligase Have Overlapping yet Distinct Roles in Phenylpropanoid Metabolism. Plant Physiology 169:2409 LP – 2421. https://doi.org/10.1104/pp.15.00838

Liang Z, Chen K, Li T, et al (2017) Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nature communications 8:14261. https://doi.org/10.1038/ncomms14261

Liedl BE, Kosier T, Desborough SL (1987) HPLC isolation and nutritional value of a major tuber protein. American Potato Journal 64:545–557. https://doi.org/10.1007/BF02853753

Lisec J, Schauer N, Kopka J, et al (2006) Gas chromatography mass spectrometry–based metabolite profiling in plants. Nature Protocols 1:387–396. https://doi.org/10.1038/nprot.2006.59

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Lloyd A, Plaisier CL, Carroll D, Drews GN (2005) Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. Proceedings of the National Academy of Sciences of the United States of America 102:2232–2237. https://doi.org/10.1073/pnas.0409339102

Lockhart DJ, Lockhart DJ, Winzeler E a, Winzeler E a (2000) Genomics, gene expression and DNA arrays. Nature 405:827–836. https://doi.org/10.1038/35015701

Lommen A, Kools HJ (2012) MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware. Metabolomics 8:719–726. https://doi.org/10.1007/s11306-011-0369-1

Lopez-Maestre H, Brinza L, Marchet C, et al (2016) SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. Nucleic Acids Research 44:e148–e148. https://doi.org/10.1093/nar/gkw655

Lu Y, Chen Q, Bu Y, et al (2017) Flavonoid Accumulation Plays an Important Role in the Rust Resistance of Malus Plant Leaves. Frontiers in plant science 8:1286. https://doi.org/10.3389/fpls.2017.01286

Lutaladio N, Ortiz O, Caldiz D (2009) Sustainable potato production. Guidelines for developing countries. Food and Agriculture Organization

Ma Z, Michailides T (2005) Advances in understanding molecular mechanisms of fungicide resistance and molecular detection of resistant genotypes in phytopathogenic fungi. Crop Protection 24:853–863. https://doi.org/10.1016/j.cropro.2005.01.011

Maher MF, Nasti RA, Vollbrecht M, et al (2020) Plant gene editing through de novo induction of meristems. Nature Biotechnology 38:84–89. https://doi.org/10.1038/s41587-019-0337-2

Malinovsky F, Fangel J, Willats W (2014) The role of the cell wall in plant immunity. Frontiers in plant science. 5:178.

Malnoy M, Viola R, Jung M-H, et al (2016) DNA-Free Genetically Edited Grapevine and Apple Protoplast Using CRISPR/Cas9 Ribonucleoproteins. Frontiers in Plant Science 7: https://doi.org/10.3389/fpls.2016.01904

Maloney V (2004) Plant metabolomics. BioTeach J 2:92–99

Mao Y, Botella JR, Liu Y, Zhu J-K (2019) Gene editing in plants: progress and challenges. National Science Review 6:421–437. https://doi.org/10.1093/nsr/nwz005

Massa AN, Childs KL, Lin H, et al (2011) The transcriptome of the reference potato genome solanum tuberosum group Phureja clone DM1-3 516R44. PLoS ONE 6: https://doi.org/10.1371/journal.pone.0026801

Matern U, Grimmig B, Edward Kneusel R (2011) Plant cell wall reinforcement in the disease-resistance response: Molecular composition and regulation. Canadian Journal of Botany 73:511–517. https://doi.org/10.1139/b95-290

Melhus IE (1915) Germination and infection with the fungus of the late blight of potato (Phytophthora infestans). Agricultural Experiment Station of the University of Wisconsin

Mendoza HA, Mihovilovich EJ, Saguma F (1996) Identification of triplex (YYYy) Potato Virus Y (PVY) immune progenitors derived fromSolanum tuberosum ssp.andigena. American Potato Journal 73:13–19. https://doi.org/10.1007/BF02849300

Meng EC, Pettersen EF, Couch GS, et al (2006) Tools for integrated sequence-structure analysis

with UCSF Chimera. BMC Bioinformatics 7:339. https://doi.org/10.1186/1471-2105-7-339

Meng S, Torto-Alalibo T, Chibucos MC, et al (2009) Common processes in pathogenesis by fungal and oomycete plant pathogens, described with Gene Ontology terms. BMC Microbiology 9:S7–S7. https://doi.org/10.1186/1471-2180-9-S1-S7

Miedes E, Vanholme R, Boerjan W, Molina A (2014) The role of the secondary cell wall in plant resistance to pathogens. Frontiers in plant science 5:358. https://doi.org/10.3389/fpls.2014.00358

Miki D, Zhang W, Zeng W, et al (2018) CRISPR/Cas9-mediated gene targeting in Arabidopsis using sequential transformation. Nature Communications 9:1967. https://doi.org/10.1038/s41467-018-04416-0

Miller J, Nolte P, Olsen N, et al (2006) Late Blight Management Action Plan for Potatoes. University of Idaho Extension http://www.cals.uidaho.edu/edcomm/pdf/CIS/CIS1132.pdf

Mills RE, Luttig CT, Larkins CE, et al (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Research 16:1182–1190. https://doi.org/10.1101/gr.4565806

Mimee B, Dauphinais N, Bélair G (2015) Life Cycle of the Golden Cyst Nematode, Globodera rostochiensis, in Quebec, Canada. Journal of Nematology 47:290–295

Molina I, Li-Beisson Y, Beisson F, et al (2009) Identification of an Arabidopsis Feruloyl-Coenzyme A Transferase Required for Suberin Synthesis. Plant Physiology 151:1317 LP – 1328. https://doi.org/10.1104/pp.109.144907

Moscou MJ, Bogdanove AJ (2009) A Simple Cipher Governs DNA Recognition by TAL Effectors. Science 326:1501–1501. https://doi.org/10.1126/science.1178817

Mosquera T, Alvarez MF, Jiménez-Gómez JM, (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. https://doi.org/10.1371/journal.pone.0156254

Muroi A, Ishihara A, Tanaka C, (2009) Accumulation of hydroxycinnamic acid amides induced

by pathogen infection and identification of agmatine coumaroyltransferase in Arabidopsis thaliana. Planta 230:517. https://doi.org/10.1007/s00425-009-0960-0

Nærstad R, Hermansen A, Bjor T (2007) Exploiting host resistance to reduce the use of fungicides to control potato late blight. Plant Pathology 56:156–166. https://doi.org/10.1111/j.1365-3059.2006.01491.x

Nakabayashi R, Yonekura-Sakakibara K, Urano K, (2014) Enhancement of oxidative and drought tolerance in Arabidopsis by overaccumulation of antioxidant flavonoids. The Plant journal : for cell and molecular biology 77:367–379. https://doi.org/10.1111/tpj.12388

Nakane E, Kawakita K, Doke N, Yoshioka H (2003) Elicitation of primary and secondary metabolism during defense in the potato. Journal of General Plant Pathology 69:378–384. https://doi.org/10.1007/s10327-003-0075-6

Nazarian-Firouzabadi F, Joshi S, Xue H, Kushalappa AC (2019) Genome-wide in silico identification of LysM-RLK genes in potato (Solanum tuberosum L.). Molecular Biology Reports 46:5005–5017. https://doi.org/10.1007/s11033-019-04951-z

Nekrasov V, Staskawicz B, Weigel D, (2013) Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nature Biotechnology 31:691–693. https://doi.org/10.1038/nbt.2655

Nelson R, Wiesner-Hanks T, Wisser R, Balint-Kurti P (2018) Navigating complexity to breed disease-resistant crops. Nature Reviews Genetics 19:21–33. https://doi.org/10.1038/nrg.2017.82

Nicot N, Hausman J-F, Hoffmann L, Evers D (2005) Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress. Journal of Experimental Botany 56:2907–2914. https://doi.org/10.1093/jxb/eri285

Niederhauser JS, Cervantes J, Servin L (1954) Late blight in Mexico and its implications.

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nature reviews Genetics 12:443–451. https://doi.org/10.1038/nrg2986

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 Disease 96:4–17. https://doi.org/10.1094/PDIS-05-11-0458

Okazaki Y, Saito K (2012) Recent advances of metabolomics in plant biotechnology. Plant Biotechnology Reports 6:1–15. https://doi.org/10.1007/s11816-011-0191-2

Oliver KM, Russell JA, Moran NA, Hunter MS (2003) Facultative bacterial symbionts in aphids confer resistance to parasitic wasps. Proceedings of the National Academy of Sciences 100:1803 LP – 1807. https://doi.org/10.1073/pnas.0335320100

Oshlack A, Robinson MD, Young MD (2010) From RNA-seq reads to differential expression results. Genome Biology 11:220. https://doi.org/10.1186/gb-2010-11-12-220

Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nature Reviews Genetics 5:389

Peng A, Chen S, Lei T, (2017) Engineering canker-resistant plants through CRISPR/Cas9-targeted editing of the susceptibility gene CsLOB1 promoter in citrus. Plant Biotechnology Journal 15:1509–1519. https://doi.org/10.1111/pbi.12733

Petolino JF (2015) Genome editing in plants via designed zinc finger nucleases. In Vitro Cellular and Developmental Biology - Plant 51

Piasecka A, Jedrzejczak-Rey N, Bednarek P (2015) Secondary metabolites in plant innate immunity: conserved function of divergent chemicals. New Phytologist 206:948–964. https://doi.org/10.1111/nph.13325

Pina-Martins F, Vieira BM, Seabra SG, (2016) 4Pipe4 – A 454 data analysis pipeline for SNP detection in datasets with no reference sequence or strain information. BMC Bioinformatics 17:41. https://doi.org/10.1186/s12859-016-0892-1

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. https://doi.org/10.1186/1471-2105-11-395

Puchta H, Baker MD, Read LR, (1999) Double-strand break-induced recombination between

ectopic homologous sequences in somatic plant cells. Genetics 152:1173–81. https://doi.org/10.1105/tpc.8.11.2057

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. Proceedings of the National Academy of Sciences 93:5055–5060

Pushpa D, Yogendra KN, Gunnaiah R, (2013) Identification of Late Blight Resistance-Related Metabolites and Genes in Potato through Nontargeted Metabolomics. Plant Molecular Biology Reporter 32:584–595. https://doi.org/10.1007/s11105-013-0665-1

Raffaele S, Win J, Cano LM, Kamoun S (2010) Analyses of genome architecture and gene expression reveal novel candidate virulence factors in the secretome of Phytophthora infestans. BMC Genomics 11:637. https://doi.org/10.1186/1471-2164-11-637

Ramakrishna G, Kaur P, Nigam D, (2018) Genome-wide identification and characterization of InDels and SNPs in Glycine max and Glycine soja for contrasting seed permeability traits. BMC plant biology 18:141. https://doi.org/10.1186/s12870-018-1341-2

Rautengarten C, Ebert B, Ouellet M, (2012) Arabidopsis Deficient in Cutin Ferulate encodes a transferase required for feruloylation of ω-hydroxy fatty acids in cutin polyester. Plant physiology 158:654–665. https://doi.org/10.1104/pp.111.187187

Rayson S, Arciga-Reyes L, Wootton L, (2012) A Role for Nonsense-Mediated mRNA Decay in Plants: Pathogen Responses Are Induced in Arabidopsis thaliana NMD Mutants. PLOS ONE 7:e31917

Reem NT, Pogorelko G, Lionetti V, (2016) Decreased Polysaccharide Feruloylation Compromises Plant Cell Wall Integrity and Increases Susceptibility to Necrotrophic Fungal Pathogens. Frontiers in plant science 7:630. https://doi.org/10.3389/fpls.2016.00630

Rey MEC, Ndunguru J, Berrie LC, (2012) Diversity of Dicotyledenous-Infecting Geminiviruses and Their Associated DNA Molecules in Southern Africa, Including the South-West Indian Ocean Islands. Viruses 4:1753–1791. https://doi.org/10.3390/v4091753

Rinaldo AR, Ayliffe M (2015) Gene targeting and editing in crop plants: a new era of precision opportunities. Molecular Breeding 35:1–15

Rogier O, Chateigner A, Amanzougarene S, (2018) Accuracy of RNAseq based SNP discovery and genotyping in Populusnigra. BMC genomics 19:909. https://doi.org/10.1186/s12864-018-5239-z

Rosenfeld JA, Ballif BC, Torchia BS, (2010) Copy number variations associated with autism spectrum disorders contribute to a spectrum of neurodevelopmental disorders. Genetics In Medicine 12:694

Salmond GPC (1992) Bacterial diseases of potatoes: from classical phytobacteriology to molecular pathogenicity. Netherlands Journal of Plant Pathology 98:115–126. https://doi.org/10.1007/BF01974478

Schauer N, Fernie AR (2006) Plant metabolomics: towards biological function and mechanism. Trends in Plant Science 11:508–516

Schindele A, Dorn A, Puchta H (2020) CRISPR/Cas brings plant biology and breeding into the fast lane. Current Opinion in Biotechnology 61:7–14. https://doi.org/https://doi.org/10.1016/j.copbio.2019.08.006

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nat Protoc 3:. https://doi.org/10.1038/nprot.2008.73

Schrider DR, Stevens K, Cardeño CM, (2011) Genome-wide analysis of retrogene polymorphisms in Drosophila melanogaster. Genome Research 21:2087–2095. https://doi.org/10.1101/gr.116434.110

Schumann GL, D'Arcy CJ (2000) Late blight of potato and tomato. The Plant Health Instructor 1–10. https://doi.org/10.1094/PHI-I-2000-0724-01

Sevestre F, Facon M, Wattebled F, Szydlowski N (2020) Facilitating gene editing in potato: a Single-Nucleotide Polymorphism (SNP) map of the Solanum tuberosum L. cv. Desiree genome. Scientific Reports 10:2045. https://doi.org/10.1038/s41598-020-58985-6

Shah SA, Erdmann S, Mojica FJM, Garrett RA (2013) Protospacer recognition motifs. RNA Biology 10:891–899. https://doi.org/10.4161/rna.23764

Shan Q, Wang Y, Li J, Gao C (2014) Genome editing in rice and wheat using the CRISPR/Cas

system. Nature protocols 9:2395–410. https://doi.org/10.1038/nprot.2014.157

Shastry BS (2009) SNPs: Impact on Gene Function and Phenotype. In: Komar AA (ed) Single Nucleotide Polymorphisms: Methods and Protocols. Humana Press, Totowa, NJ, pp 3–22

Shaul O (2015) Unique Aspects of Plant Nonsense-Mediated mRNA Decay. Trends in Plant Science 20:767–779. https://doi.org/https://doi.org/10.1016/j.tplants.2015.08.011

Shulaev V (2006) Metabolomics technology and bioinformatics. Briefings in Bioinformatics 7:128–139. https://doi.org/10.1093/bib/bbl012

Singh J, Kaur L (2016) Chemistry, Processing, and Nutritional Attributes of Potatoes—An Introduction BT - Advances in Potato Chemistry and Technology (Second Edition). Academic Press, San Diego, pp xxiii–xxvi

Singh J, Kaur L, McCarthy OJ, (2009) Development and characterization of extruded snacks from New Zealand Taewa (Maori potato) flours. Food Research International 42:666–673. https://doi.org/10.1016/j.foodres.2009.02.012

Smith RA, Cass CL, Mazaheri M, (2017) Suppression of CINNAMOYL-CoA REDUCTASE increases the level of monolignol ferulates incorporated into maize lignins. Biotechnology for Biofuels 10:109. https://doi.org/10.1186/s13068-017-0793-1

Solovyev V (2007) Statistical Approaches in Eukaryotic Gene Prediction. Handbook of Statistical Genetics 97–159

STATCAN (2016)**.** Statistics Canada. Canadian Potato Production.

Steinert J, Schiml S, Puchta H (2016) Homology-based double-strand break-induced genome engineering in plants. Plant Cell Reports 35:1429–1438

Svitashev S, Schwartz C, Lenderts B, (2016) Genome editing in maize directed by CRISPR–Cas9 ribonucleoprotein complexes. Nature Communications 7:13274

Symington L, Gautier J (2011a) Double-strand break end resection and repair pathway choice. Annual review of genetics

Symington LS, Gautier J (2011b) Double-Strand Break End Resection and Repair Pathway

Choice. Annual Review of Genetics 45:247–271. https://doi.org/10.1146/annurev-genet-110410-132435

Tagousop CN, Tamokou J-D, Ekom SE, (2018) Antimicrobial activities of flavonoid glycosides from Graptophyllum grandulosum and their mechanism of antibacterial action. BMC Complementary and Alternative Medicine 18:252. https://doi.org/10.1186/s12906-018-2321-7

Tahira T, Kukita Y, Higasa K, (2009) Estimation of SNP Allele Frequencies by SSCP Analysis of Pooled DNA. In: Komar AA (ed) Single Nucleotide Polymorphisms: Methods and Protocols. Humana Press, Totowa, NJ, pp 193–207

Tohge T, Fernie AR (2010) Combining genetic diversity, informatics and metabolomics to facilitate annotation of plant gene function. Nature Protocols 5:1210–1227. https://doi.org/10.1038/nprot.2010.82

Trapnell C, Roberts A, Goff L, (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7:562–78. https://doi.org/10.1038/nprot.2012.016

Tsuda K, Katagiri F (2010) Comparing signaling mechanisms engaged in pattern-triggered and effector-triggered immunity. Current Opinion in Plant Biology 13:459–465. https://doi.org/10.1016/j.pbi.2010.04.006

Uhse S, Djamei A (2018) Effectors of plant-colonizing fungi and beyond. PLoS pathogens 14:e1006992–e1006992. https://doi.org/10.1371/journal.ppat.1006992

Uitdewilligen JGAML, Wolters A-MA, D'hoop BB, (2013) A Next-Generation Sequencing Method for Genotyping-by-Sequencing of Highly Heterozygous Autotetraploid Potato. PLOS ONE 8:e62355

van Esse HP, Reuber TL, van der Does D (2020) Genetic modification to improve disease resistance in crops. New Phytologist 225:70–86. https://doi.org/10.1111/nph.15967

Vanholme R, Cesarino I, Rataj K, (2013) Caffeoyl Shikimate Esterase (CSE) Is an Enzyme in the Lignin Biosynthetic Pathway in Arabidopsis Science 341:1103 LP – 1106. https://doi.org/10.1126/science.1241602

Vaser R, Adusumalli S, Leng SN, (2015) SIFT missense predictions for genomes. Nature Protocols 11:1

Veillet F, Chauvin L, Kermarrec M-P, (2019a) The Solanum tuberosum GBSSI gene: a target for assessing gene and base editing in tetraploid potato. Plant Cell Reports 38:1065–1080. https://doi.org/10.1007/s00299-019-02426-w

Veillet F, Perrot L, Chauvin L, (2019b) Transgene-Free Genome Editing in Tomato and Potato Plants Using Agrobacterium-Mediated Delivery of a CRISPR/Cas9 Cytidine Base Editor. International Journal of Molecular Sciences 20

Vidal RO, do Nascimento LC, Mondego JMC, (2012) Identification of SNPs in RNA-seq data of two cultivars of Glycine max (soybean) differing in drought resistance. Genetics and molecular biology 35:331–334. https://doi.org/10.1590/S1415-47572012000200014

Vleeshouwers VGAA, van Dooijeweert W, Govers F, (2000) The hypersensitive response is associated with host and nonhost resistance to Phytophthora infestans. Planta 210:853–864. https://doi.org/10.1007/s004250050690

Vossen JH, van Arkel G, Bergervoet M, (2016) The Solanum demissum R8 late blight resistance gene is an Sw-5 homologue that has been deployed worldwide in late blight resistant varieties. TAG Theoretical and applied genetics Theoretische und angewandte Genetik 129:1785–1796. https://doi.org/10.1007/s00122-016-2740-0

Wang C-H, Yu J, Cai Y-X, (2016) Characterization and Functional Analysis of 4-Coumarate:CoA Ligase Genes in Mulberry. PLOS ONE 11:e0155814

Wang H, Fan W, Li H, (2013a) Functional characterization of Dihydroflavonol-4-reductase in anthocyanin biosynthesis of purple sweet potato underlies the direct evidence of anthocyanins function against abiotic stresses. PloS one 8:e78484–e78484. https://doi.org/10.1371/journal.pone.0078484

Wang M, Lu Y, Botella JR, (2017) Gene Targeting by Homology-Directed Repair in Rice Using a Geminivirus-Based CRISPR/Cas9 System. Molecular Plant 10:1007–1010

Wang S, Zhang S, Wang W, (2015) Efficient targeted mutagenesis in potato by the CRISPR/Cas9 system. Plant Cell Reports 34:1473–1476. https://doi.org/10.1007/s00299-015-1816-7

Wang Y, Chantreau M, Sibout R, Hawkins S (2013b) Plant cell wall lignification and monolignol metabolism. Frontiers in plant science 4:220. https://doi.org/10.3389/fpls.2013.00220

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews Genetics 10:57–63. https://doi.org/10.1038/nrg2484

Watanabe K (2015) Potato genetics, genomics, and applications. Breeding Science 65:53–68. https://doi.org/10.1270/jsbbs.65.53

Wen W, Li D, Li X, (2014) Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. Nature Communications 5:3438. https://doi.org/10.1038/ncomms4438

Were HK, Kabira JN, Kinyua ZM, (2013) Occurrence and Distribution of Potato Pests and Diseases in Kenya. Potato Research 56:325–342. https://doi.org/10.1007/s11540-013-9246-9

Wiesner-Hanks T, Nelson R (2016) Multiple Disease Resistance in Plants

Woo JW, Kim J, Kwon S Il, (2015) DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. Nature Biotechnology 33:1162–1164. https://doi.org/10.1038/nbt.3389

Wu J, Kang J-H, Hettenhausen C, Baldwin IT (2007) Nonsense-mediated mRNA decay (NMD) silences the accumulation of aberrant trypsin proteinase inhibitor mRNA in Nicotiana attenuata. The Plant Journal 51:693–706. https://doi.org/10.1111/j.1365-313X.2007.03173.x

Xie K, Yang Y (2013) RNA-Guided genome editing in plants using a CRISPR-Cas system. Molecular Plant 6:1975–1983. https://doi.org/10.1093/mp/sst119

Xu X, Pan S, Cheng S, (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195. https://doi.org/10.1038/nature10158

Yang C, Liang Y, Qiu D, (2018a) Lignin metabolism involves Botrytis cinerea BcGs1- induced defense response in tomato. BMC Plant Biology 18:103. https://doi.org/10.1186/s12870-018-1319-0

Yang J, Yan R, Roy A, (2015) The I-TASSER Suite: protein structure and function prediction.

Nature methods 12:7–8. https://doi.org/10.1038/nmeth.3213

Yang Q, He Y, Kabahuma M, (2017) A gene encoding maize caffeoyl-CoA O-methyltransferase confers quantitative resistance to multiple pathogens. Nature Genetics 49:1364

Yang Q, Xuan Trinh H, Imai S, (2004) Analysis of the Involvement of Hydroxyanthranilate Hydroxycinnamoyltransferase and Caffeoyl-CoA 3-O-Methyltransferase in Phytoalexin Biosynthesis in Oat. Molecular Plant-Microbe Interactions® 17:81–89. https://doi.org/10.1094/MPMI.2004.17.1.81

Yang X, Guo X, Yang Y, (2018b) Gene Profiling in Late Blight Resistance in Potato Genotype SD20. International journal of molecular sciences 19:1728. https://doi.org/10.3390/ijms19061728

Yin K, Gao C, Qiu J-L (2017) Progress and prospects in plant genome editing. Nature Plants 3:17107

Yogendra KN, Kumar A, Sarkar K, (2015a) Transcription factor StWRKY1 regulates phenylpropanoid metabolites conferring late blight resistance in potato. Journal of experimental botany 66:7377–89. https://doi.org/10.1093/jxb/erv434

Yogendra KN, Kushalappa AC (2016) Integrated transcriptomics and metabolomics reveal induction of hierarchies of resistance genes in potato against late blight. Functional Plant Biology 15:497–7828. https://doi.org/10.1071/FP16028

Yogendra KN, Kushalappa AC, Sarmiento F, (2015b) Metabolomics deciphers quantitative resistance mechanisms in diploid potato clones against late blight. Functional Plant Biology 42:284–298. https://doi.org/10.1071/FP14177

Yogendra KN, Pushpa D, Mosa KA, (2014a) Quantitative resistance in potato leaves to late blight associated with induced hydroxycinnamic acid amides. Functional and Integrative Genomics 14:285–298. https://doi.org/10.1007/s10142-013-0358-8

Yogendra KN, Pushpa D, Mosa KA, (2014b) Quantitative resistance in potato leaves to late blight associated with induced hydroxycinnamic acid amides. Functional & integrative genomics 14:285–98. https://doi.org/10.1007/s10142-013-0358-8

Yoshida K, Schuenemann VJ, Cano LM, (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. eLife 2:e00731. https://doi.org/10.7554/eLife.00731

You KT, Li LS, Kim N-G, (2007) Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. PLoS biology 5:e109–e109. https://doi.org/10.1371/journal.pbio.0050109

Yuan J, Zhu M, Lightfoot DA, (2008) In silico comparison of transcript abundances during Arabidopsis thaliana and Glycine max resistance to Fusarium virguliforme. BMC Genomics 9:S6–S6. https://doi.org/10.1186/1471-2164-9-S2-S6

Zaidi SS-A, Mansoor S (2017) Viral Vectors for Plant Genome Engineering. Frontiers in Plant Science 8:. https://doi.org/10.3389/fpls.2017.00539

Zhang B, Yang X, Yang C, (2016a) Exploiting the CRISPR/Cas9 System for Targeted Genome Mutagenesis in Petunia. Scientific Reports 6:20315

Zhang C, Wang P, Tang D, (2019) The genetic basis of inbreeding depression in potato. Nature Genetics 51:374–378. https://doi.org/10.1038/s41588-018-0319-1

Zhang F, Maeder ML, Unger-Wallace E, (2010) High frequency targeted mutagenesis in Arabidopsis thaliana using zinc finger nucleases. Proceedings of the National Academy of Sciences of the United States of America 107:12028–12033. https://doi.org/10.1073/pnas.0914991107

Zhang F, Voytas DF (2011) Targeted mutagenesis in Arabidopsis using zinc-finger nucleases. Methods in molecular biology (Clifton, NJ) 701:167–177. https://doi.org/10.1007/978-1-61737-957-4_9

Zhang R, Miner JJ, Gorman MJ, (2016b) A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. Nature 535:164–168. https://doi.org/10.1038/nature18625

Zhang Y, Zhang F, Li X, (2013) Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. Plant Physiology 161:20–27. https://doi.org/10.1104/pp.112.205179

Zhao Q, Dixon RA (2014) Altering the Cell Wall and Its Impact on Plant Disease: From Forage to Bioenergy. Annual Review of Phytopathology 52:69–91. https://doi.org/10.1146/annurev-phyto-082712-102237

Zhao Y, Wang K, Wang W, (2019) A high-throughput SNP discovery strategy for RNA-seq data. BMC Genomics 20:160. https://doi.org/10.1186/s12864-019-5533-4

Zhao Y, Zhang C, Liu W, (2016) An alternative strategy for targeted gene replacement in plants using a dual-sgRNA/Cas9 design. Scientific Reports 6:23890. https://doi.org/10.1038/srep23890

Zheng Q, Cai X, Tan MH, (2014) Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. BioTechniques 57:115–124. https://doi.org/10.2144/000114196

APPENDIX

Appendix 4.1: Stepwise CRISPR vector construction: sequences and primers used.

Target sequences selected:

T1: GTGTTGTTCCCATGTAAAAATCA

T2: CTTTGAGAGGCTCTGGTTCTCGA

T5: TGCAAATCTCCCTGATTTTTACA

T6: GCAAATCTCCCTGATTTTTACAT

> **Arabidopsis U6-1 promoter sequence**

AGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTCGATCCGTAGAAACGAGAC
GGTCATTGTTTTAGTTCCACCACGATTATATTTGAAATTTACGTGAGTGTGAGTGAG
ACTTGCATAAGAAAATAAAATCTTTAGTTGGGAAAAAATTCAATAATATAAATGGG
CTTGAGAAGGAAGCGAGGGATAGGCCTTTTTCTAAAATAGGCCCATTTAAGCTATTA
ACAATCTTCAAAAGTACCACAGCGCTTAGGTAAAGAAAGCAGCTGAGTTTATATAT
GGTTAGAGACGAAGTAGTGATT (304 bp)

>**gRNA scaffold sequence**

GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAA
AGTGGCACCGAGTCGGTGC (76 bp)

TTTTTTT (7 bp)

CTAGACCCAGCTTTCTTGTACAAAGTTGGCATTA (34 bp)

> *S.tuberosum* DNA for U6 small nuclear RNA promoter region

> **Z17290.1** S.tuberosum DNA for U6 small nuclear RNA promoter region
AAATGGTACAAGTTGAATATGGGGGCAAATCTGGACTCTAGGCTTAGTTGGGCTCTA
TGTGCATATAAAAGCAAGAGCAAAAACCTGTAGCTAGGTCCAGGCCCATGCCTTTG
GTAAAACTCAATGTGCTAATTCTCCCTCATCGTCTGCAGAGAGAAGCCTCGCTGTGT
TTATATAATTGAACAGTAACATGCATGCTT

Multiple guide RNA (gRNA) expression cassettes and cloning into puc119-MCS

gRNA_*StCCoAOMT*_ cassette 1 (sgRNA1):

Contains Potato U6 promoter drive guideRNA (targeting *StCCoAOMT* gene target site 2): (334 bp)

CGCGGA**AAGCTT****AAATGGTACAAGTTGAATATGGGGGCAAATCTGGACTCTAG**
GCTTAGTTGGGCTCTATGTGCATATAAAAGCAAGAGCAAAAACCTGTAGCTAG
GTCCAGGCCCATGCCTTTGGTAAAACTCAATGTGCTAATTCTCCCTCATCGTCT

GCAGAGAGAAGCCTCGCTGTGTTTATATAATTGAACAGTAACATGCATGCTTGC
TTTGAGAGGCTCTGGTTCTCGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAA
GGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT**GGT**
ACCTCTAGCTAG

Blue – Restriction sites (*HindIII* and *KpnI*)
Red - Potato U6 promoter sequence
Black - target site 2 (gRNA1)
Green - gRNA-Cas9 scaffold

Primers:
StCaoCOMT-gRNA1-F: CGCGGAAAGCTTAAATGGTACAAGTTGAAT
StCaoCOMT-gRNA1-R: CTAGCTAGAGGTACCAAAAAAAGCACCGAC

gRNA_*StCaoCOMT*_ cassette 2 (sgRNA2):

Contains Arabidopsis U6 promoter drive guideRNA (targeting *StCCoAOMT* gene target site 6): (438 bp)

GATCGATCTG**CTGCAG**AGAAATCTCAAAATTCCGGCAGAACAATTTTGAATCTC
GATCCGTAGAAACGAGACGGTCATTGTTTTAGTTCCACCACGATTATATTTGAA
ATTTACGTGAGTGTGAGTGAGACTTGCATAAGAAAATAAAATCTTTAGTTGGGA
AAAAATTCAATAATATAAATGGGCTTGAGAAGGAAGCGAGGGATAGGCCTTTT
TCTAAAATAGGCCCATTTAAGCTATTAACAATCTTCAAAAGTACCACAGCGCTT
AGGTAAAGAAAGCAGCTGAGTTTATATATGGTTAGAGACGAAGTAGTGATTGC
AAATCTCCCTGATTTTTACATGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAG
GCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTTGAGC
TCAGGCGC

Blue – Restriction sites (*PstI* and *SacI*)
Red - Potato U6 promoter sequence
Black - target site 6 (gRNA2)
Green - gRNA-Cas9 scaffold

Primers:
StCCoAOMT-gRNA2-F: GATCGATCTGCTGCAGAGAAATCTCAAAAT
StCCoAOMT-gRNA2-R: GCGCCTGAGCTCAAAAAAAGCACCGACTCG

Primers for confirmation puc119-MCS:

Verify the sequence accuracy of the cloned gRNA expression cassette by Sanger sequencing. Sequencing primer F (sequencing from *EcoRI* toward *HindIII* in pUC119 - MCS) is 5′ ATTAAGTTGG GTAACGCC 3′ and primer R (sequencing from *HindIII* toward *EcoRI*) is 5′ TGGAATTGTGAGCGGATA 3′.

pUC119-MCS-F: ATTAAGTTGGGTAACGCC

pUC119-MCS-R: TGGAATTGTGAGCGGATA

pFGC-pco-cas9 vector sequencing primer:

pFGC-pco-cas9_F: AATAAAAACTGACTCGGA