Investigating wheat rust virulence evolution through transcriptome analysis of a recently emerged race of *Puccinia triticina*

A Thesis Submitted to the Committee on Graduate Studies in Partial Fulfillment of the Requirements for the Degree of Master of Science in the Faculty of Arts and Science

TRENT UNIVERSITY
Peterborough, Ontario, Canada

©Copyright by Kayla Margaret Marsh, 2017

Environmental and Life Sciences M.Sc. Graduate Program
January 2017
ABSTRACT

Investigating wheat rust virulence evolution through transcriptome analysis of a recently emerged race of *Puccinia triticina*

Kayla Margaret Marsh

*Puccinia triticina*, wheat leaf rust (WLR), is the most economically damaging fungal rust of wheat on a global scale. This study identified transcriptome changes in a recently emerged race of WLR in Ontario with a new virulence type relative to a possible ancestor race. Also, this study focused on detecting variation in candidate virulence genes and uncovering novel insight into WLR virulence evolution. Various race-by-variety interactions were evaluated using RNA-seq experiments. A list of genes with statistically significant expression changes in each comparison was prepared and predicted effectors were retained for further analysis. Proteins with nonsynonymous substitutions were run through BLASTx to identify potential orthologs. Over 100 candidate effectors with a 2-fold or higher change in transcript level were identified. Seven of these candidate effector genes were recognized to contain single nucleotide polymorphisms (SNPs) which altered the amino acid sequence of the resulting protein. The information gained may aid in targeted breeding programs to combat new WLR races as well as provide the basis for functional analysis of WLR using potential orthologs in a model basidiomycete.

**Keywords:** rust fungi, wheat leaf rust, WLR, transcriptome, RNA-seq, effectors, SNPs, virulence
ACKNOWLEDGEMENTS

The last few years have been the most challenging, both for personal reasons as well as academic. I have had so many ups and downs while pushing through to completion of this degree, and I have had so many people who have supported me in many different ways. I am so thankful for the support and love of everyone around me.

First and foremost, I would like to thank my supervisor Dr. Barry Saville. His constant and unwavering support, both academically and personally has helped me get to where I am and helped me become who I am today. I appreciate the chance I had to complete this project and grow as a researcher and scientist. I would also like to thank the members of my supervisory committee, Dr. Christopher Kyle and Dr. Marcel Dorken for their feedback and encouragement during completion of my thesis.

I would like to thank the Grain Farmers of Ontario for the funding and providing me the opportunity to do exciting research. Without the funding, this project would not have been possible. As well I would like to thank AAFC and those within the company that directly helped train and teach me, and were always available to answer questions.

I would also like to send a great big thank you to the past and present grad students and research technicians of the Saville Laboratory. You have been such a pleasure to know and work with and I would not be where I am today without your support, knowledge and friendship.
Thank you to Matthew, who, unfortunately, got to deal with me while I was knee-deep in this project. Thank you for being patient, understanding, caring and carrying us when I didn’t have much to offer.

Finally, I need to thank my friends and family, who have been an emotional pillar during these last few years. Each and every one of you, my parents, my siblings, my aunt and grandma, I love you all and appreciate your support, love and laughs. Specifically, my mom and brothers, who would travel so far to see me and spend even just a day with me, you do not know how much I appreciate that and needed it to stay strong.

Thank you everyone. This is dedicated to you all.
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................................................... ii
ACKNOWLEDGEMENTS ...................................................................................................................... iii
TABLE OF CONTENTS ......................................................................................................................... v
LIST OF TABLES ................................................................................................................................. viii
LIST OF ABBREVIATIONS ................................................................................................................... ix

CHAPTER 1: INTRODUCTION .............................................................................................................. 1
General Background ............................................................................................................................... 1
Breeding for Resistance ......................................................................................................................... 4
Effectors .............................................................................................................................................. 6
Gene–for-gene Theory ........................................................................................................................... 8
Lifecycle of WLR ................................................................................................................................. 9
Wheat Leaf Rust in Ontario .................................................................................................................. 14

CHAPTER 2: LITERATURE REVIEW .................................................................................................... 19
Relevance .......................................................................................................................................... 19
Introduction to Effectors ................................................................................................................... 19
Bacterial Effector Proteins ............................................................................................................... 21
Oomycete Effector Proteins ............................................................................................................. 23
Fungal Effector Proteins ................................................................................................................... 25
The Plant Immune System .................................................................................................................. 28
The Gene-for-Gene Theory ................................................................................................................. 31
Indirect Interactions – The Guard Theory ......................................................................................... 34
Enzymatic Activities of Effectors ..................................................................................................... 35
Conclusions ....................................................................................................................................... 36

CHAPTER 3: METHODS ....................................................................................................................... 37
Plant Growth ...................................................................................................................................... 37
Seedling Inoculation ........................................................................................................................... 38
Sample Collection .............................................................................................................................. 40
RNA Isolation .................................................................................................................................... 41
Samples from Clinical Genomics Centre ........................................................................................... 46
Samples from British Columbia Genome Sequencing Centre (BCGSC) ........................................ 46
RNA-seq Analysis .................................................................................................................. 48
Identifying Secretion Signals ................................................................................................. 49
Single Nucleotide Polymorphism (SNP) Identification ........................................................... 55

CHAPTER 4: RESULTS ........................................................................................................... 60
Tissue Growth and Sample Collection .................................................................................. 60
RNA Isolation .......................................................................................................................... 61
RNA-Seq Quality Control ..................................................................................................... 64
Transcriptome Sequencing and Mapping .............................................................................. 66
Normalization of Experiment Data ........................................................................................ 67
Expression Analysis and Identification of Candidate Effector Genes .................................... 69
Investigation of Amino Acid Changes .................................................................................... 72

CHAPTER 5: DISCUSSION ..................................................................................................... 76
Transcriptome Sequencing and Mapping .............................................................................. 76
Expression Analysis and Identification of Candidate Effector Genes .................................... 78
In Silico Protein Analysis ...................................................................................................... 81
Future Directions .................................................................................................................. 83

CHAPTER 6: CONCLUSIONS ............................................................................................... 87
REFERENCES ......................................................................................................................... 89
APPENDICES ........................................................................................................................ 100
LIST OF FIGURES

Figure 1: Nomenclature system developed for Wheat Leaf Rust.................................14
Figure 2: Life cycle of Puccinia triticina taken from Bolton, 2008.........................24
Figure 3: Virulence patterns on various leaf rust resistance (Lr) genes in Ontario........27
Figure 4: Layout of planted wheat seeds..................................................................48
Figure 5: Custom sprayer attachment........................................................................50
Figure 6: Example of an output from ProtComp v 9.0............................................62
Figure 7: Example of an output from SignalP 4.1.....................................................64
Figure 8: Example of a TargetP 1.1 output..............................................................65
Figure 9: In silico comparisons of WLR races..........................................................68
Figure 10: Flow chart depicting the analyses on the RNA-seq data..........................70
Figure 11: Infection phenotypes of WLR.................................................................72
Figure 12: 1.5% BPTE agarose gel depicting quality of RNA samples....................74
Figure 13: An example of normalization of expression data.....................................79
Figure 14: A volcano plot of expression.................................................................80
**LIST OF TABLES**

Table 1: Rust race and host combinations used in infection time courses……………….52

Table 2: AmpliTaq master mix………………………………………………………………56

Table 3: NanoDrop Spectrophotometer output………………………………………………73

Table 4: Quality Control (QC) results for samples submitted for RNA-seq………………76

Table 5: Mapping statistics of paired reads for each sample……………………………..77

Table 6: Potential effectors and those with SNPs resulting in changes to the amino acid sequences………………………………………………………………………………...82

Table 7: Nonsynonymous SNPs present in candidate effector genes…………………..84

Table 8: Potential orthologs to the putative effector proteins…………………………..86
LIST OF ABBREVIATIONS

°C  degrees Celsius
ADP  adenosine diphosphate
ATP  adenosine triphosphate
AVR  avirulence gene/protein
bp  basepair
CC  coiled-coil
cm  centimeter(s)
cTP  chloroplastic protein
DEPC  diethylpyrocarbonate
DNA  deoxyribonucleic acid
DPI  days post infection
ER  endoplasmic reticulum
EST  expressed sequence tag
ETI  effector triggered immunity
g  gram(s)
gDNA  genomic deoxyribonucleic acid
HMM  Hidden Markov Model
HR  hypersensitive response
mTP  mitochondrial protein
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Loc</td>
<td>localization prediction</td>
</tr>
<tr>
<td>Lr</td>
<td>leaf rust resistance gene</td>
</tr>
<tr>
<td>LRR</td>
<td>leucine-rich repeat</td>
</tr>
<tr>
<td>Mb</td>
<td>megabases</td>
</tr>
<tr>
<td>mg</td>
<td>milligram(s)</td>
</tr>
<tr>
<td>mL</td>
<td>millilitre(s)</td>
</tr>
<tr>
<td>mM</td>
<td>millimolar</td>
</tr>
<tr>
<td>NBS</td>
<td>nucleotide binding site</td>
</tr>
<tr>
<td>ng</td>
<td>nanogram(s)</td>
</tr>
<tr>
<td>NLR</td>
<td>nucleotide binding/leucine-rich repeat</td>
</tr>
<tr>
<td>nm</td>
<td>nanometers</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAMP</td>
<td>pathogen-associated molecular pattern(s)</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PRR</td>
<td>PAMP-recognition receptor(s)</td>
</tr>
<tr>
<td>psi</td>
<td>pounds per square inch</td>
</tr>
<tr>
<td>PTI</td>
<td>PAMP-triggered immunity</td>
</tr>
<tr>
<td>QC</td>
<td>quality control</td>
</tr>
<tr>
<td>R</td>
<td>resistance gene/protein</td>
</tr>
<tr>
<td>RC</td>
<td>reliability class</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>RPKM</td>
<td>reads per kilobase of transcript per million mapped reads</td>
</tr>
<tr>
<td>RQI</td>
<td>RNA quality index</td>
</tr>
<tr>
<td>SAR</td>
<td>systemic acquired resistance</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>SSP</td>
<td>small, secreted protein</td>
</tr>
<tr>
<td>ssRNA</td>
<td>single stranded ribonucleic acid</td>
</tr>
<tr>
<td>SUMO</td>
<td>small, ubiquitin-like modifiers</td>
</tr>
<tr>
<td>T3E</td>
<td>type III effector</td>
</tr>
<tr>
<td>TTSS</td>
<td>type III secretion system</td>
</tr>
<tr>
<td>WLR</td>
<td>wheat leaf rust</td>
</tr>
<tr>
<td>µM</td>
<td>micromolar</td>
</tr>
<tr>
<td>µg</td>
<td>microgram(s)</td>
</tr>
<tr>
<td>µL</td>
<td>microlitre(s)</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

General Background

Throughout the history of agriculture, cereal crop production has been threatened by fungal diseases. Of these diseases, biotrophic rusts of wheat have been the most economically damaging (Hodson, 2011). Severe epidemics caused by rust periodically occur in all regions where wheat (*Triticum aestivum*) is grown (Roelfs *et al.*, 1992). Rust fungi are recorded to have been connected to epidemics and devastation of crops by Aristotle, and rust urediniospores found in Israel have been dated to 1300 BC (Kislev, 1982). The causal agent of stem rust of wheat, *Puccinia graminis*, was named in 1797 (Chester, 1946). At that time, wheat leaf rust was not distinguished from stem rust, but by 1815 leaf rust was described as a distinct fungus (Chester, 1946; Roelfs *et al.*, 1992). Wheat leaf rust, *Puccinia triticina*, is the most common and widely distributed of all cereal rusts, occurring wherever wheat is grown (Roelfs, 1989). Wheat cultivars that are susceptible to leaf rust regularly suffer yield losses of 5-15% or more depending on the stage of development of the wheat at the time of initial rust infection (Kolmer, 1996; Samborski DJ, 1985). The amount of loss due to fungal infection has two dependencies: the stage of crop development at the time of spore germination and initial infection and the resistance type bred into the particular wheat cultivar (Chester, 1946; Bolton, 2008). Lower percentage of loss will be experienced when initial infection of the fungus occurs on mature crops. This is due, in part, to the age of the wheat and how many times the fungus may cycle and re-infect (Chester, 1946). Further, to reduce loss associated with infection, the resistance type bred into the wheat cultivar must be the same type as the
rust race which is infecting in order for recognition of pathogen and subsequent defense responses to take place.

The virulence of wheat rust fungi against wheat resistance genes is highly diverse, resulting in many different pathogenic races being present in wheat fields around the world. Specific interactions between resistance genes present in wheat cultivars and avirulence genes in the rust serve as useful markers for characterizing rust populations. These markers are used in annual surveys of wheat leaf rust virulence phenotypes that are conducted in both Canada (McCallum and Seto-Goh, 2004) and the USA (Kolmer et al., 2004). The markers are identified by inoculating the rust races on a series of near-isogenic lines of wheat (differentials), which differ only by the presence of a single rust resistance gene. The ability to grow in a specific manner on these lines identifies the genotype of the rust races. A four-letter nomenclature system has been developed for the rust fungi and race designations are determined by testing the infection type of the fungus on these differentials, seen in Figure 1 (Long and Kolmer, 1989; McCallum, Seto-Goh and Xue, 2010). Populations can be characterized by distribution of races and the frequencies of virulence against specific rust resistance genes on this defined set of differential wheat hosts (Kolmer, 2005).
Figure 1: Nomenclature system developed for Wheat Leaf Rust. L= low infection type (resistant) H= high infection type. Four sets of differentials are tested at a time (sets denoted in blue) and given a letter (green column) depending on the infection type seen (middle of table). Sets 1-3 are outlined in Long and Kolmer, 1989 and Set 4 is outlined in McCallum et al., 2010.

Breeding of genetic resistance to reduce yield losses due to rust fungi is economical when compared to other options such as fungicide. However, wheat breeding programs throughout the world have mixed results in producing durable cultivars with long-lasting and effective resistance to leaf rust (Kolmer, 1996). Knowledge of the identity of resistance genes in commonly used parental germplasm and released cultivars is crucial for the breeding of successful genetic resistance. Identification of leaf rust resistance genes allows for the incorporation of different genes into germplasm pools.
which helps to breed genetically robust cultivars in relation to current rust endemics (Kolmer, 1994).

**Breeding for Resistance**

WLR has been a prevalent disease on wheat for thousands of years and is more damaging now than ever as wheat cultivars, which are the same or closely related genetically, are grown contiguously over large areas of land. This landscape provides an easy target for the infectious spores of WLR (Samborski, 1985; Bolton, 2008; Webb and Fellers, 2006). A major part in the defense against loss due to rust fungi is breeding modern wheat varieties to include a genetic basis for disease resistance. This practice began in the early 1900’s when the genetics of breeding, in particular for disease resistance, was beginning to be understood (Biffin, 1905). Genetic resistance reduces losses from WLR by incorporating previously identified leaf rust resistance (Lr) genes into the wheat cultivar. These bred-in resistances are effective in reducing the yield losses of wheat for approximately six years before losing effectiveness due to imposed selection for virulent rust races (Kilpatrick, 1975). The breeding process to produce new wheat cultivars can require as many as 10 years to complete (McIntosh and Brown, 1997). Thus, the ability to predict possible upcoming mutations and pathotypes of rust is of utmost importance. Due to the lack of the secondary host in most of the growing area of Ontario, mutation in the pathogen, often documented as a single mutation or first in stepwise mutations, is required to generate virulence in the fungi on a cultivar that was previously resistant (McIntosh and Brown, 1997). Pre-emptive crossing of cultivars and anticipatory selection for resistance is performed and once a new virulence pathotype is detected,
replacement cultivars which are resistant can be incorporated almost at once into farmers’ fields since they are already bred (McIntosh and Brown, 1997). Resistant wheat varieties being implemented into fields quickly help to diminish overall infection of wheat growing areas by individually staving off infection.

When plants are attacked by pathogens, they implement two main lines of defense. The primary system is the basal defense against potential pathogens, which relies on the recognition of conserved microbial features. These features are often referred to as pathogen-associated molecular patterns (PAMPs) which are recognized by PAMP-recognition receptors (PRRs). This recognition activates PAMP-triggered immunity (PTI) and this immunity prevents further colonization of the host plant (De Wit, 2007). The second line of defense is active after the basal line is overcome. This is a more specialized recognition system directed by effector perception by resistance (R) proteins and the subsequent activation of an effector-triggered immunity (ETI). This immunity leads to a rapid and acute response, the hallmark of which is the hypersensitive response (HR) (Jones, 2006). It is this second line of defense which triggers the coevolutionary arms race seen between plants and plant pathogens. This arm race consists of pathogens mutating so that proteins may better avoid or suppress the ETI and the plant reacts by developing novel R proteins to enhance the recognition of the novel proteins secreted by the pathogen.
**Effectors**

During infection, biotrophic fungi develop feeding relationships with their hosts by intimately associating with the plant cells. To do this, specialized structures are used within the infected host tissues, such as the hyphae which grows through the plant apoplast, or haustoria which invaginates the host’s plasma membrane. The hyphae and haustoria were historically described as feeding structures, pulling nutrients from the host, but more recently these have emerged also as sites of secretion and translocation into the host cells for pathogen virulence proteins, known as effectors (Raffaele *et al.*, 2010).

All proteins and small molecules secreted by the pathogen that alter the host cell structure or function to facilitate infection or to trigger defense responses in the host are classified as effectors. The protein effectors can manipulate plant processes to the advantage of the fungus by promoting host infection and suppressing plant defense responses. Since a hallmark of many effectors is that they are released from the fungus and enter the host cells, an initial step in identifying this class of effectors is to first identify open reading frames (ORFs) with predicted signal peptides for protein secretion and a lack of internal membrane-spanning domains.

Effectors are a subset of the “small secreted” peptide (SSP) group, which the majority of are under 300 amino acids in length (Duplessis *et al.* 2011) and include all proteins and small molecules secreted by the pathogen that alter the host cell structure to facilitate infection in the host. A major subclass of effectors are proteins which are cysteine rich and contain intramolecular disulfide bridges. These bridges stabilize tertiary structure of the protein within the harsh environments found in the host plant, such as the
apoplast. Effectors from oomycetes were among the first to be studied. These effectors were often found to contain conserved host-translocation motifs for transport into the host plant cytoplasm from the fungus. Though oomycetes have this conservation, no universal motif has been identified for fungi. Further, effectors are often evolutionarily diverse and novel, and are rarely similar to already characterized proteins, therefore the search for effectors cannot be limited to sequence similarity (Saunders et al., 2011). What effectors from different organisms do have in common is that they are secreted from the attacking microbe into the host in some way.

How or where effectors are secreted dictates how the effector is classified, which is in one of three ways: 1) through the fungal endoplasmic reticulum (ER) secretory pathway and into the fungus-host interface, 2) apoplastic effectors which stay in the plant extracellular space, or 3) cytoplasmic effectors which enter the host cell and target different cellular components in the host (Rafiqi et al., 2012). The biological function as well as the mechanism of delivery of these effector proteins is mostly unknown due to the large number of diverse and novel effectors, as well as the potential for functional redundancy among them. Further, the lack of functional signatures in the amino acid sequences, the species-specific specialization and lack of methods available for genetic manipulation for obligate biotrophic fungal pathogens make it inherently difficult to expand the understanding of them (Rafiqi et al., 2012).

As mentioned, fungal effector proteins are classified in three ways, but of those they can be roughly categorized into two groups, extracellular effectors which are secreted into the apoplast of the host plant and cytoplasmic effectors which are translocated into the host cells. Extracellular effectors are often N-terminally processed
by plant or fungal proteases, though they can sometimes be C-terminally processed. Due to the obligate nature of biotrophic fungi, the evidence for this protein maturation is based in some cases solely on *in-silico* predictions. Besides the processing, another common feature of extracellular effectors is the presence of multiple cysteine residues which, as mentioned are involved in the disulfide-bridge formation which provides stability to the proteins within the high protease environment of the host apoplast (Stergiopoulos and deWit, 2009).

**Gene–for-gene Theory**

As mentioned, plants have developed two main lines of defense against infection which have co-evolved with fungal infectors. These lines of defense include mechanisms which are either physical or chemical barriers to prevent initial infection or the spread thereof. Host plant responses to infection involve rapid gene activation and networks of signal transduction (Yang *et al.*, 1997). Study of the coevolution of plant hosts and pathogens have resulted in the discovery of a gene-for-gene relationship. Resistance and susceptibility to plant disease are dictated by the combined genotypes of the plant host and the pathogen. The potential for resistance in the host plant and the pathogen’s ability to cause disease are both controlled by pairs of matching genes; in the host plant, a resistance (*R*) gene and in the fungus, an avirulence (*Avr*) gene. It is these *Avr* proteins which are a subset of effectors from the fungus. However, plant hosts producing the protein associated with the *R* gene are resistant to a pathogen which produces corresponding to the *Avr* gene product (Flor, 1942) rendering these ‘effectors’ incapable of infection in that specific host and therefore they are referred to as *Avr* proteins instead.
Others describe this theory as a plant having one locus with two alleles for resistance and susceptibility and a pathogen having a corresponding locus with two alleles for virulence and avirulence (Brown and Tellier, 2011). The interaction of these products depends on signals and responses involved during initial stages of attack (Yang et al., 1997).

The main difference between a resistant and susceptible interaction is the immediate recognition of the invading pathogen. Resistant or incompatible reactions occur when a protein translated from the resistance gene in the host plant interacts with a specific protein from an avirulence gene in the pathogen. The interaction between the products of these genes leads to the activation of host defense responses, such as the hypersensitive response, that stops the growth of the fungus (Flor, 1942). Susceptible or compatible reactions occur in the absence of an avirulence protein in the pathogen and are the result of slow and weak responses that fail to restrict the colonization and spread of the pathogen, often due to the fact that an avirulent pathogen is not recognized by the host plant or a susceptible plant does not recognize the pathogen (Yang et al., 1997; Kolmer, 2005; Brown and Tellier, 2011). The proteins in this reaction, instead of being known as Avr proteins, are often referred to as effectors, since they influence infection and further proliferation of the fungus.

**Lifecycle of WLR**

*Puccinia triticina* is a macrocyclic and heteroecious fungus with five spore stages and two unrelated host species, which are required to complete the full lifecycle. On the surface of an infected wheat leaf (*Triticum aestivum* L.), dikaryotic urediniospores are
produced in pustules known as uredinia (Figure 2A). The uredial stage of the lifecycle can result in infection of the host in humid and warm conditions within roughly 10 days (Bolton, 2008). As the spores mature, they continue to re-infect the wheat leaves (Bolton, 2008). As well as re-infecting the current host, the urediniospores can be wind dispersed over thousands of kilometers, initiating infection of distant wheat plants if conditions are favourable (Szabo and Kolmer, 2007). The urediniospores germinate soon after making contact with the epidermis of the wheat leaf, resulting in the creation of a germ tube. This tube extends along the epidermis of the leaf until it reaches a stomatal pore. Appressorium are formed at the stoma which facilitate fungal penetration of the plant (Hu et al., 2007). After penetration, the germ tube differentiates into a substomatal vesicle followed by a haustorial mother cell. This mother cell is responsible for interactions such as breaching the cell wall and forming an intercellular haustorium which establishes nourishment of the fungus via the host plant (Hu et al., 2007). It is at this point in the infection where the differences between some compatible and resistant interactions can be seen. A virulent interaction results in further branching of the haustorial mother cell through the plant tissue, producing intercellular hyphae and haustoria. A resistant reaction typically halts with the haustorial mother cell, as the plant defense system becomes active (Hu et al., 2007). The urediniospores re-infect the host until the leaves are too old, at that point the next stage in the fungal lifecycle begins.

At leaf senescence, brown-black, dikaryotic teliospores are produced within structures known as telia, found beneath the leaf epidermis and approximately the same size as uredinia (Figure 2B). Early in development of this spore, the haploid nuclei undergo karyogamy to produce diploid nuclei. The teliospores germinate to produce a
promycelium, eventually giving rise to haploid basidiospores (Figure 2B; bottom inset). To do this, the diploid teliospores complete meiosis to create haploid nuclei and these migrate into the forming basidiospores. Each basidiospore receives one haploid nucleus, which undergoes mitosis to create a mature basidiospore with two haploid nuclei. Of the four basidiospores found on the basidium, there are two of each mating type (+ and -).

Basidiospores are ejected from the basidium and wind-dispersed onto a nearby alternative host, meadow rue (Thalictrum speciosissimum). However, T. speciosissimum is rarely found in most wheat growing areas in North America thus this section of the lifecycle is often skipped, relying on the re-infection of urediniospores (Bolton, 2008; Webb et al., 2006). When the basidiospores find meadow rue, they infect the epidermal leaf cells of the plant, leading to the production of pustules called pycinia (Figure 2C).

Within the pycnia, haploid pycniospores and receptive hyphae are formed, which act as the male and female gametes, respectively (Figure 2D). Due to the heterothallic nature of the fungus, pycniospores and hyphae originating from the same pycnium are not sexually compatible and spores from different pycnium must be disseminated to surrounding pycnia containing sexually compatible hyphae in order to complete fertilization. To accomplish this, the pycnium produces a nectar type liquid surrounding the pycniospores. This nectar attracts insects to aid in spore movement around the leaf surface. The fungus also relies on dew movement and rain splashing on the leaf surface to disperse spores.

The fusion of compatible mating pairs (+/-) results in fertilization, which restores the dikaryotic nuclear state in the resulting mycelium. These mycelium proliferate through the leaf and lead to the development of a dikaryotic aecium on the underside of the leaf, typically directly underneath the pycnium (Figure 2E). Chains of dikaryotic aeciospores
are produced from the aecium. Once these spores erupt through the epidermis of the leaf, they are wind-disseminated to the uredinial host (wheat). The aeciospores germinate on the wheat leaf surface and penetrate stomata, resulting in the production of asexual dikaryotic urediniospores. WLR can cycle indefinitely as asexual uredinal infections (Roelfs, 1989).
Figure 2: Life cycle of *Puccinia triticina* taken from Bolton, 2008. (A) Uredinia on wheat leaf containing single cell dikaryotic urediniospores. Top inset: surface view of a uredinium (×100); bottom inset: echinulate surface of a single urediniospore (×3000). (B) Telia on the leaf epidermis. Top inset: teliospores which originate from telia; bottom inset: Promycelium with four haploid basidiospores. (C) Pycnia on *Thalictrum*. (D) Diagrammatic cross-section of *Thalictrum* with pycnia. Top inset: basidiospores infect *Thalictrum* in which the fungus produces haploid pycnia (×400); bottom inset: pycniospores and flexuous hyphae are immersed in a liquid ‘nectar’ exudate (×500). (E) Diagrammatic cross-section of *Thalictrum* with both pycnia and aecia. Top inset: following fertilization, a dikaryotic aecium develops (×200); bottom inset: the aecium produces chains of dikaryotic aeciospores (×1250).
As stated, the alternate host, *T. speciosissimum*, is rarely found in the wheat growing areas in Canada and the United States, which supports the theory that WLR is primarily cycling as asexual infections with urediniospores and does not have the opportunity to complete the sexual part of its lifecycle (Webb et al., 2006; Saari et al., 1968). This is very important as it indicates that any changes seen in the DNA sequence are more likely to be linked with mutation and single nucleotide polymorphisms (SNPs) than with sexual recombination. Further, any changes to pathotypes and infection capability would be a result of those mutations. In North America, new pathotypes are constantly emerging, even with the absence of the secondary host. In Ontario specifically, the emergence of new pathotypes is tracked by members of Agriculture and Agri-Food Canada (AAFC) as well as the Ontario Ministry of Agriculture, Food and Rural Affairs (OMAFRA).

**Wheat Leaf Rust in Ontario**

Races of WLR present in Ontario are tracked annually to maintain a database indicating which pathotypes are currently infecting wheat plants. This database also aids in predicting which races may become a problem in the future. The near-isogenic differentials, as mentioned previously, are used to genotype emerging WLR races. The differentials enable researchers to distinguish races of wheat leaf rust.

Scoring WLR races using this panel of wheat differentials allowed McCallum and collaborators to genotype the wheat leaf rust races in Ontario fields and infer that a new leaf rust race has emerged. This specific race has overcome the resistance provided by the
gene \textit{Lr24}, present in the popular wheat variety Vienna, shown in Figure 3 (McCallum and Seto-Goh, 2009). Figure 3A depicts the dramatic increase in virulence on wheat with \textit{Lr24} seen in Ontario between the years 2004 and 2005, with a further increase in 2006. The rise of this virulence has been attributed to the substantial increase in the wheat leaf rust race MFDS in Ontario in 2006, consisting of 73.7\% of leaf rust infections (Figure 3B). In previous years, the dominant rust genotype seen in Canadian wheat fields was MBDS (Figure 3C). It was proposed that MFDS and MBDS are closely related rust races which is interesting since MBDS and MFDS levels found are somewhat reversely related, where MBDS decreases as MFDS levels increases (Figure 3C).
Figure 3: Virulence patterns on various leaf rust resistance (Lr) genes in Ontario (McCallum and Seto-Goh, 2009). (A) Percent of isolates of four pathotypes over the years 1999 to 2006 in Ontario. Outlined is a dramatic increase in Lr24. (B) The number of isolates found as well as the percent of all isolates for two different pathotypes, MFDS and TDBJ, in Saskatchewan and Manitoba (SK + MB), Ontario (ON), Quebec (QC), Prince Edward Island (PEI) and overall in Canada for the year 2006. Outlined is the rise of MFDS in Ontario. (C) Percentage of various virulence phenotypes found in Canadian wheat fields over multiple years. Outlined is the decrease of type MBDS and increase of type MFDS.
It was hypothesized by my collaborator Brent McCallum, of Agriculture and Agri-Food Canada, that the emerged wheat leaf rust race with genotype MFDS, which is virulent on \textit{Lr24}, arose from the rust race MBDS, avirulent on \textit{Lr24}. Since the secondary plant host, \textit{Thalictrum speciosissimum}, is rarely present in wheat growing areas of North America, the sexual reproductive cycle of WLR cannot be completed and sexual recombination is likely not a large contributing factor to genotype changes. Instead, a mutation or series of mutations or SNPs at the DNA level is/are likely the cause.

In order to identify changes, a transcriptome analysis was performed on RNA-seq data. RNA-seq data will allow both for interpreting the sequence of the DNA and how it may have been altered in terms of SNPs when compared to a reference sequence, as well as identify specifically which genes are being expressed during infection. The approach of this analysis was threefold; 1) to identify rust genes whose expression differs during infection by races MBDS and MFDS, 2) to identify candidate virulence genes, and 3) to provide insight regarding the molecular basis of virulence change in rust fungi.

RNA-seq provides a huge wealth of data. Within these data, a large number of differences in the transcriptomes of each interaction was expected. In starting the analysis, specific infection profiles were compared to each other. Generally, these comparisons were used to determine potential virulence gene influence on gene expression of the fungus during infection in each of various interactions. To narrow the pool of candidate genes to those potentially involved in virulence or as being effector proteins, I completed several \textit{in silico} subtractions between these compared infection profiles. Further, I identified those gene encoding proteins that may be secreted and have no known orthologs, which may be potential effector proteins. Lastly, within those
candidate effectors, I searched for mutations when compared with a reference sequence and determined the impact that these mutations may have on the translated proteins.
CHAPTER 2: LITERATURE REVIEW

Relevance

This thesis describes mining for data within an RNA-seq database created for this project from infections by different races of *P. triticina* on varying wheat varieties. Specifically, this project aimed to find differentially expressed genes whose translated proteins are small, potentially novel and with sequences that indicate they are secreted from the fungus. These are identified in this project as potential effector proteins. Further, these differentially expressed genes were also searched for the presence of SNPs identified by comparing the same genes from different infection profiles. This literature review outlines a historical view of plant pathogen effectors, what they are and how they are known to function. There is a great deal of research done on some organisms, such as those we consider to be models, but there is very little done in this regard with the rust fungi. It is important to understand how these effectors may be functioning in related species so that this knowledge may be applied when searching for effectors within the rust species.

Introduction to Effectors

Host plants have evolved multiple layers of resistance to help protect against invasion from bacterial, oomycete and fungal pathogens and increase the survival of plant cells from these pathogens. To combat this, pathogens have evolved systems which compensate. It is thought that secreted proteins, known as effectors, are the key to
successful host infection by fungal, bacterial and oomycete pathogens (Rafiqi et al., 2012). These effector proteins are virulence factors encoded by genes in the pathogen which are secreted during host invasion and facilitate suppression of the plant defense response system and alter the host to make it a more hospitable environment for the fungus. In some cases, these proteins aid to alter infected tissue so that it becomes a source of nutrients for the invading pathogen which will support its further growth and development (Koeck et al., 2011).

Effector research primarily focuses on bacteria and oomycete species and some research is conducted on fungal ascomycete effectors, including Venturia inaequalis the causal agent of Apple scab disease (Bowen et al., 2009). Basidiomycetes, which include the rust species, have limited information regarding effectors at the current time, compared to the knowledge we have on bacterial effectors. Most of the effectors researched are known to be small, secreted, cysteine-rich proteins which have a direct or indirect primary role in virulence and a secondary role to inhibit plant defenses allowing the further penetration and expansion of infection (Stergiopoulos and deWit, 2009; Dénes et al., 2015). The majority of these molecules are typically under 300 amino acids in length (Duplessis et al. 2011), though, a few effectors can be much larger and still effectively function as such, an example being the Ustilago maydis virulence factor Cmu1 (Djamei et al., 2011). The following sections of this paper will discuss the more well-researched areas of effector proteins, such as bacterial and oomycete effectors, and outline research in the less characterized fungal effectors.
**Bacterial Effector Proteins**

Bacterial pathogens have evolved a variety of mechanisms to increase pathogenicity which include modification, addition or deletion of genes to increase the pathogen’s ability to adapt to both differing environments and circumstances. The mechanisms by which effector proteins enter the plant cell cytoplasm from bacteria are very well-characterized, in contrast to those in fungi and oomycetes (Petre and Kamoun, 2014). Identification of existing bacterial effector proteins, which mimic host proteins involved in signal transduction pathways, have led to insights into bacterial pathogenesis as well as host mimicry employed by bacterial proteins. These processes are used by the pathogen to interfere with host signaling and signal transduction processes (Whisson et al., 2007). All major groups of gram-negative bacterial pathogens use a specific adaptation known as the type III secretion system (TTSS) which uses a complex structure to directly translocate bacterial effector proteins into host cells (Galán and Wolf-Watz, 2006).

Secretion via the TTSS is a two-step process involving the secretion of the proteins across the cytoplasmic membrane followed by a second secretion across the outer membrane into the extracellular area. These effectors associated with the TTSS are often referred to as type three effectors (T3E). The TTSS is important to pathogenicity of certain bacterial genera that colonize the apoplastic of plants and typically elicit cell death in the plant, including *Pseudomonas* and *Pantoea* (Alfano and Collmer, 2004). The TTSS pathway is encoded by *hrp* and *hrc* genes. The proteins of these genes aid in the secretion of TTSS substrates across the bacterial envelope and then the translocation of effectors through the host cell respectively (Cornelis and Van Gijsegem, 2000).
As well as enabling the translocation of proteins, the TTSS allows for proteins to enter an otherwise impenetrable host. When investigating the regulation of the cytoskeleton of eukaryotes, the ExoS protein of the pathogen *Pseudomonas aeruginosa* was found to have two distinct domains, causing the protein to be bifunctional. Both of these domains are thought to disrupt or alter the cytoskeleton, with the N-terminus itself of the ExoS protein being sufficient by disrupting actin in hamster ovary cells in order to alter the cytoskeleton (Steele-Mortimer *et al.*, 2002; Pederson *et al.*, 1999). Thus, the ExoS protein is one of the effectors which are the basis for by-passing the cytoskeleton and allowing for infection.

There are multiple strategies used by T3Es from bacteria to minimize response from the host and increase pathogen infection. The main strategy identified is receptor targeting. Effectors may target specific receptors involved in plant immunity response and either degrade or inhibit them from performing their intended action. The AvrPtoB effector from *Pseudomonas syringae* pv. *glycinea*, the causal agent of bacterial blight of soybean, uses a ligase domain to induce a proteasome-mediated degradation of immunity response proteins (Gimenez-Ibanez *et al.*, 2009). A specific example of an immunity response protein which is targeted by AvrPtoB is the FLS2 protein in *P. syringae*. AvrPto, also found in *P. syringae*, targets specific proteins, such as EFR (receptor for elongation factor Tu (EF-Tu)), in order to inhibit their activity and block triggered immunity in the host (Shan *et al.*, 2008). EF-Tu elicits the defense response to pathogen attacks. AvrPto blocks the defense response by blocking the receptor for EF-Tu. AvrPto also interacts with the BAK1 protein and is able to inhibit it from interacting with FLS2, the complex of which is also involved in immune response of the host (Shan *et al.*, 2008).
The BAK1 protein itself may be a conserved target for many T3Es from different bacterial species that infect multiple plant hosts, including a *Xanthomonas oryzae pv oryzae* (Xoo) T3E, Xoo2875, which interacts with the rice BAK1 ortholog (Yamaguchi *et al.*, 2013).

The type III secretion system is not the only system found to be in use in bacteria, though it is the most commonly used. Other bacterial effectors use different systems, such as the type VI secretion system used to compete against other bacteria or the type II secretion system which delivers a vast array of virulence determinants directly to the plant apoplast (Hood *et al.*, 2010; Jha *et al.*, 2005).

**Oomycete Effector Proteins**

Similar to the type III system above, oomycetes introduce their effectors into host plant apoplast and cytoplasm, potentially through haustoria, to enable molecular reprogramming of the host defense responses. A few Avr oomycete genes have been identified within multiple species including the *ATR13* and *ATR1NdWsB* genes from the Arabidopsis pathogen *Hyloperonospora parasitica* (Allen *et al.*, 2004). These genes are detected by specific R proteins within the cytoplasm of the plant host. The gene products contain a conserved RXLR motif within the signal peptides (Rehmany *et al.*, 2005). This motif has been found to be similar to the RXLX(E/Q) motif in malaria which is a host-cell targeting signal required for translocating proteins into host erythrocytes (Marti *et al.*, 2004). During infection with malaria parasite, parasitophorous vacuole (PV) structures are created to protect the parasites within the host while the parasite replicates. The N-
terminal signal peptides of the parasite enable the secretion of effectors into the PV and it is the host-cell targeting signal which allows the translocation of the effector proteins through the PV membrane and into the red blood cells of the host (Przyborski and Lanzer, 2004).

Effectors which are delivered through haustoria may use RXLR motifs to direct translocation of those effectors into the plant cells. This RXLR motif is common among the species _Phytophthora infestans_, _Phytophthora sojae_, and _Phytophthora ramorum_ and is present in at least 100 proteins predicted to be secreted from these organisms. These findings support the notion that the motif is involved in effector translocation (Birch et al., 2006). Furthermore, it was found in _P. infestans_ that an RXLR-containing protein also contains a functional nuclear localization signal and this particular protein may accumulate in the nuclei of the host during infection (Birch et al., 2006).

Important progress in the study of oomycete effectors has been made leading to the identification of large repertoires of effectors with characteristic RXLR and other motifs potentially required for host cell uptake. The mechanisms which are employed by effectors in suppressing host immunity to increase virulence have been investigated by different groups (Bos et al., 2010; Mukhtar et al., 2011; Wang et al., 2011).
Fungal Effector Proteins

Fungal effectors are being increasingly studied, largely due to the enormous yield losses in agricultural crops caused by fungal pathogens, with rice blast, grey mould and the wheat rusts at the forefront (Dean et al., 2012). These crop losses are easily translated to economic, environmental and ecological losses as well as impacting human health which is likely the main reason these diseases are so well studied.

Obligate biotrophic fungal pathogens and their effectors are inherently difficult to study. Researchers can overcome this by employing the use of bacterial type III secretion systems for delivering foreign proteins into problematic systems of study. This system was developed in 2011 with the intent to find a system to deliver effectors into wheat without causing a resistance response. The process began with testing within a bacterial setting. Adding a TTSS to Pseudomonas fluorescens enabled the successful delivery of bacterial effector proteins AvrRpm1 and AvrRpt2 while not inducing an immediate resistance response in the host (Yin and Hulbert, 2011). This system was also successfully used in 2014 using the engineered P. fluorescens to introduce the AvrRpm1 protein from P. syringae and AvrBs2 protein from X. campestris into a wheat host (Upadhyaya et al., 2014). Identifying that this system can be used with certain effectors to successfully infiltrate a wheat host without inducing a resistance response gives hope that this tool may be used in the future to shuttle fungal effectors into a host in a similar manner for study.

Fungal effector proteins are classified as two main groups: those which are extracellular and secreted into the apoplast or xylem of the host plant, and those which are cytoplasmic and are translocated into the cells of the host (Stergiopoulos and deWit,
There are also a small number of fungal effectors which are not secreted from the fungus, as evidenced in a few ascomycetes. The Ace1 protein from the *Magnaporthe grisea*, rice blast fungus, is one such protein which is instead involved in the biosynthesis of a secondary metabolite which is potentially secreted (Böhnert *et al.* 2004). Most effectors code for small secreted proteins, some of which are translocated into plant host cells (Stergiopoulos and deWit, 2009). Extracellular effectors are often processed further, either N- or C- terminally, by proteases from either the plant or the fungal pathogen. This is seen in the Tomato leaf mold pathogen, *Cladosporium fulvum*, effector proteins Avr and Ecp (deWit *et al.*, 2009). This is also seen in the Avr-Pita effector from *M. grisea* which is predicted to be both a secreted and a processed protein. Concluding that these effectors were acted upon by proteases was, however, based on the fact that the smaller mature protein and not the slightly larger intact protein is the active form which interacts directly with a complementary protein in the host plant (Jia *et al.*, 2000).

Extracellular effectors, specifically those from ascomycetes and oomycetes, also have multiple cysteine residues in their structure, which may be involved in the formation of disulfide-bridges, providing stability to the protein when in the high protease environment of the host apoplast (Stergiopoulos and deWit, 2009). Both Avr4 and Avr9 proteins of *C. fulvum* have been shown to have cysteine residues which are required for both activity and stability (Van Den Burg *et al.*, 2003; Van Den Burg *et al.*, 2001). Further members of this group are SCR74 and SCR91 proteins of the oomycete *P. infestans*, which are secreted in the early stage of infection. These proteins activate phenylalanine ammonia-lyase which is an enzyme involved in secondary metabolite production and contributes to leaf whitening (Liu *et al.*, 2004; Kamoun, 2006; Orsomando
et al., 2011). However, not all cysteine residues are involved in formation of disulfide bridges. Mutational analysis of the cysteine residues in the Ecps protein of C. fulvum showed that there were not involved in the formation of disulfide bridges but were still needed to induce a defense response in the host plant (Luderer et al., 2002).

Biotrophic fungi form haustoria within the living plant cells during infection of the host. The haustoria are essential for nutrient acquisition by the fungus, which may include manipulation of the host metabolism and/or the suppression of host defense mechanisms (Voegele and Mendgen, 2003). Of the biotrophic fungi, the basidiomycete rust and smut fungi have substantial information regarding their effectors. With the exception of the flax rust fungus, Melampsora lini, rust fungi have been found to have an exceptionally large set of secreted proteins within their proteome, many of which are not functionally annotated. The rust fungi, aside from M. lini, infect multiple hosts in order to complete their lifecycles, and may have different effector groups for each host (Libera et al., 2015). Within M. lini, Avr gene families have been identified: AvrL567, AvrM, AvrP123 and AvrP4 (Catanzariti et al., 2006). All of these encode small secreted proteins expressed within the haustoria (Kobayashi et al., 1997). Potential effector gene families and clusters are also found within the corn smut fungus, Ustilago maydis. There are over 400 identified secreted proteins within the U. maydis genome, of which about 20% reside in clusters. There are 12 clusters containing anywhere from three to 26 genes which have been identified to be regulated together through expression analysis. Deletion in 5 of these clusters have been shown to alter virulence of the fungus, indicating a role in infection (Kämper et al., 2006). Among the secreted proteins of U. maydis is Pep1, which supresses plant immunity by inhibiting POX12, a component in the reactive
oxygen species (ROS)-generating system, and Pit2, which inhibits cysteine proteases of the host involved in salicylic acid-associated defenses (Hemetsberger et al., 2012; Mueller et al., 2013).

There are a few different avenues by which effectors can function in order to improve the virulence of the pathogen within the host. Some effector proteins work as virulence factors themselves, being required for infection of the host plant. There are also effector genes that are located within clusters of genes related to pathogenesis which are correlated with virulence on certain hosts. An example of such are the genes Six1, Six2, and Six3 in Fusarium oxysporum which are clustered in a section of the genome known to be correlated with virulence on tomato plants (Van Der Does et al., 2008). There are some effector proteins which work in combination with other effectors to maintain virulence of the pathogen. This may mean that their individual involvement in virulence is undetectable or redundant. Finally, there are effector proteins which provide either overlapping effects or general roles that include interference with defense signaling pathways in the host plant (Stergiopoulos and deWit, 2009).

**The Plant Immune System**

Plants are potential hosts for a large group of diverse pathogens such as fungi, oomycetes and bacteria among others due to their immobile nature. However, plants are able to mount defenses to invading pathogens both locally and systemically, even while lacking a circulatory system (Fu and Dong, 2013). The cells of plants are protected by layers of physical barriers, the first being the waxy surface of the leaf then the cell wall
and finally the plasma membrane within the cell. All of these barriers to the cells aid the plant itself in denying access to invading microbes. Along with the physical barriers, plants have certain chemical barriers to invasion, such as saponins which disrupt cell membranes of fungal pathogens using their soap-like properties (Bednarek and Osbourn, 2009). During early stages of pathogen colonization on a host plant, adapted pathogens secrete effector molecules that alter immune responses and facilitate colonization by the pathogen. The host plant can reinstate immunity to the pathogen through evolution of the receptors which recognize these new effectors. It is this recognition which elicits responses in the plant to cut off the invading pathogen (Thomma et al., 2011). Plants have evolved two lines of active defense in response to pathogen attack. The first is a basal defense against pathogens which involves surface pattern recognition receptors (PRRs) which, when recognizing a pathogen-associated molecular pattern (PAMP), induce a response in the host plant called a pattern-triggered immunity (PTI) (Zipfel, 2014). It is this PTI which prevents further colonization of the plant. In response, pathogens are able to adapt to host genotypes and defense mechanisms through the evolution of virulence factors known as effectors. Some effectors are secreted into the host and interfere with the PTI from within host cells to reduce these basal defenses. Deslandes and Rivas, 2012 lists examples of bacterial effector proteins in current literature which act to interfere and suppress the PTI. It is likely this interference by the effectors in the host which drive the coevolution between the host plant and the pathogen by continuing cycles of plant defense and pathogen counter defense, often referred to as the evolutionary ‘arms race’.

The defense system of the plant employs polymorphic intracellular nucleotide-binding/leucine-rich repeat (NLR) receptors to intercept invading effectors and
effectively induce a reaction called the effector-triggered immunity (ETI) to stop pathogen growth (Jacob et al., 2013). This ETI is a secondary defense to the basal PTI and is often associated with localized plant cell death which is known as the hypersensitive response (HR) (Jones and Dangl, 2006). The different N-terminal domains of the NLR proteins fall into two categories, those with a coiled-coil (CC) domain, known as CNLs, and those with a Toll-interleukin 1 receptor (TIR) domain, known as TNLs (Cui et al., 2015). NLR receptors have evolved differently in both plants and mammals, resulting in related, but structurally different nucleotide binding/oligomerization domain/leucine-rich-repeat (NOD-LRR) molecules in mammals. Besides immune system responses, these NOD-LRRs are associated with regulation of major histocompatibility complex genes involved in adaptive immunity, as well as the regulation of cell death (Ting et al., 2010). NLR activation in plants is accomplished through the N-terminal TIR or CC and the C-terminal LRR domains which cooperate sterically. This cooperation inhibits nucleotide-binding domains within ATP/ADP exchange (Griebel et al., 2014). There are mechanisms within the host plant that restrict the expression of NLR receptor genes when they are not required in order to maintain plant growth (Heidrick et al., 2012).

As mentioned, a pathogen attempting to infect a host may trigger an HR. This attempted infection, however, may also induce a systemic acquired resistance (SAR) in the plant, resulting in the uninfected tissues which surround the initial infection site to be more resilient to subsequent pathogen attack. The surrounding tissues experience an accumulation of the hormone salicylic acid, used in defense responses. This accumulation hastens the response to a subsequent attack since the hormone is already present. Further,
the tissues surrounding an initial infection site also contain antimicrobial pathogenesis-related proteins, which may last in the plant for weeks to months post initial infection (Fu and Dong, 2013).

The immune response system and defense of plant hosts against pathogens is multilayered and intricate. Components of the system, such as NLR receptor genes within plant genomes, are rapidly evolving and are a likely a key part in plant pathogen interaction. These interactions have been long studied, creating theories regarding their mechanisms of function, one such being the classic gene-for-gene theory. This particular theory pre-dates many molecular discoveries, including the discovery of NLR receptor genes, but research into the area of plant-pathogen interaction provide molecular insight into aspects of the theory.

**The Gene-for-Gene Theory**

Classically, the gene-for-gene hypothesis states that for every dominant avirulence gene ($Avr$) found in the pathogen, there is a complementary resistance gene ($R$) found in the host. It is the direct interaction between the products of these genes that lead to defense responses in the host that halt the proliferation of the fungi (Flor, 1942). To elaborate, the plant variety producing an $R$ protein is resistant towards a pathogen strain or race that produces the corresponding $Avr$ protein (Thomma et al., 2011). These $Avr$ genes can also have their subsequent protein fall into the effector category. The product of some $Avr$ genes are secreted into the host, potentially to alter the host in some way to
make it more pliable for infection. Since the proposal of this hypothesis, evidence of this concept has been found. In 1984 the first bacterial Avr gene was discovered from *Pseudomonas syringae* (Staskawicz et al., 1984). It was found that when using cosmids with specific race fragments cloned in, they were able to induce other races to mimic the incompatibility of infection of the race from which cloned DNA was obtained. The first fungal Avr gene was identified in 1991 in *Cladosporium fulvum*, the causal agent of tomato leaf mold (van Kan et al., 1991). It was hypothesized that a race-specific hypersensitive response on a specific variety of tomato genotype, Cf9, was due to the production of a specific protein. It was found that this specific protein, called Avr9, was a product of an avirulence gene which was absent in other fungal races which were virulent on the Cf9 tomato genotypes. This finding indicated that the presence of avirulence genes somehow plays a role in the resistance of the host plant. The first oomycete Avr gene was identified in 2004 from *Phytophthora sojae*, the soil-borne pathogen causing stem and root rot of soybean (Shan et al., 2004). They synthesized an avirulence protein found in the pathogen, Avr1b-1, in the *Pichia pastoris* yeast. This protein was able to infiltrate soybean leaves and trigger a defense response in those plants with the Rps1b resistance gene.

The direct interaction between Avr and R proteins has been studied and a few examples have been found. The earliest detected was through the analysis of the flax and flax rust, *Melampsora lini*, interaction which led to the gene-for-gene theory (Flor, 1956). The flax plant is resistant to rust infection when the plant carries at least one resistance gene which corresponds specifically to avirulence genes present in the rust strain currently attacking. Another example is the ascomycete *Magnaporthe oryzae*, formerly
M. grisea. In yeast, binding assays demonstrated a physical interaction between the Avr-Pi-ta protein and the Pi-ta R protein in rice. The Pi-ta gene codes for an intracellular NBS-LRR protein. The interaction between the two proteins can be hindered by mutating an amino acid in the LRR domain of the Pi-ta and this disrupts the defense responses related to it (Jia et al., 2003).

There are many organisms where the gene-for-gene interaction has not been established. For the corn smut fungus, Ustilago maydis, where this interaction is not seen, several effector genes have been studied for their function in virulence. Pep1 is a novel effector protein found in U. maydis that is essential during penetration of the plant. This protein is secreted into the plant apoplast and accumulated. It was found that disruption of the gene does not affect the development of infection structures. However, two of the four cysteine residues were found to be essential for its function (Doehlemann et al., 2009).

The Stp1 gene in U. maydis encodes a secreted effector which is essential for the initial establishment of hyphae in the epidermal layer of the plant cell (Lanver et al., 2014). When deleted, Stp1 mutants are avirulent as the fungus cannot proliferate during infection (Liang and Kommen, 2012). These are examples of effector genes which work within the basis of the gene-for-gene theory. Besides the gene-for-gene theory, there are other explanations for indirect interactions which take place between proteins that influence the resistance characteristics of a host plant/pathogen system. One such is the Guard Theory.
Indirect Interactions – The Guard Theory

An implication of the gene-for-gene hypothesis is that given the very large number of diverse effector molecules in different strains of each pathogen, plant hosts must carry equally large numbers of corresponding R proteins to be able to recognize all individual effectors. However, an approach to address this used by most plants is the production of R proteins which are able to identify modifications in the plant targets of incoming fungal effectors. These R proteins do not interact directly with an effector but guard the target protein of the host plant. If alterations to the target are detected, usually by an invading effector protein, the R protein will then initiate a defense response.

Furthermore, experiments designed to show the direct interaction between these R and Avr proteins produced mostly negative results with only two direct interactions being revealed (Jia et al., 2000; Tang et al., 1996). This lack of evidence for direct interactions between R and Avr proteins led to the formulation of a new hypothesis known as the ‘guard theory’ (Van der Biezen and Jones, 1998). In this theory, it is predicted that the R proteins interact with a third (or more) of the structure of the plant protein known as the guardee to activate resistance in the plant. This guardee is targeted specifically and may be modified by the pathogen protein. Resistance is triggered in the plant host when the R protein detects an interaction or attack taking place with its guardee protein (Van der Biezen and Jones, 1998). After the creation of this hypothesis, evidence supporting it was reported. A RIN4 protein was shown to interact with secreted molecules from Pseudomonas syringae in order to elicit a resistance response in Arabidopsis thaliana plants (Mackey et al., 2002). Two effectors, AvrB and AvrRpm1, promote phosphorylation of RIN4 which is detected by the R protein RPM1. This specific
detection by a guard protein is shared in soybean, however, in this plant, recognition of both these effectors is mediated by NLRs Rpg1b and Rpg1r, which are not orthologs to RPM1 (Ashfield et al., 1995). There are some effectors which have been found to contain certain motifs, indicating that they may have enzymatic functions within the interactions.

**Enzymatic Activities of Effectors**

There are many insights into the function of effectors based on the recognition of motifs that predict enzymatic activities and plant targets. A large portion of effectors seem to be cysteine proteases, which are enzymes that degrade proteins (Alfano and Collmer, 2004). In bacterial pathogens, these proteins are able to modify the signal transduction pathways by removing small ubiquitin-like modifiers (SUMO) from pathways involving host defense (Orth et al., 2000). The XopD protein from *Xanthomonas campestris*, acts on plant proteins that have been modified by SUMO proteins, showing a protease ability (Hotson et al., 2003). Also, the AvrRpt2 effector protein in *P syringae* is a putative cysteine protease required for the elimination of a protein related to the regulation of host basal defenses, RIN4 (Axtell et al., 2003).

These activities can be used to help categorize the effectors within TTSS. Assessing whether the effectors modify any of the components found within the transduction pathway involved in resistance, it was found that there are two categories: a modification which is reversible, or the destruction or irreversible denaturation of the component being interacted with. The XopD protein mentioned previously modifies its target by removing the SUMO portion from proteins within the plant host. This process
can be reversed using native plant proteins which reattaches the SUMO groups (Hotson et al., 2003).

**Conclusions**

Effector proteins and their impacts regarding plants resistance response and fungal virulence is currently in the spotlight for research (Dénes et al., 2015; Petre and Kamoun, 2014). There have been rapid advances in sequencing techniques, allowing for the generation of both genomic and transcriptomic data from many fungal species (Buermans and Dunnen, 2014; Kim et al., 2016). Using bioinformatic tools, which are also quickly advancing, enables researchers to discover increasing amounts of novel effectors (Lanver et al., 2014; Hemetsberger et al., 2012; Mueller et al., 2013). Further, research has enabled the study of less tractable species and interactions by introducing bacterial systems into fungi (Yin and Hulbert, 2011; Upadhyaya et al., 2014). Typically, the function of a potential effector is confirmed experimentally, using gene deletion or overexpression. Use of secretion systems within bacteria allow for the study of effectors in systems where gene manipulation is not possible.
CHAPTER 3: METHODS

Plant Growth

20 cm (8”) mum pots from ITML (Brantford, ON) were used to grow wheat plants. The bottom ¾ of the pots were filled with Sunshine soil professional growing mix #1/LC1 (Sungro Horticulture Canada) and 70 wheat seeds of one variety were placed onto this layer. The seeds were organized as laid out in Figure 4.

Figure 4: Layout of wheat seeds planted into Sunshine soil professional growing mix. X indicates location of seeds with 5 seeds being placed into each X.
This layer of seeds was covered with a mixture of a 1:1 soil (Sunshine mix #1/LC1) and fine granulated, screened sand (All Treat Farms Play Sand) to fill the remaining ¼ of the planting pot. These full pots were thoroughly watered and covered with plastic to increase the humidity in the pot and encourage seed germination. The pots, with the plastic cover, were placed into a Conviron PGC 20 growth chamber where they were incubated (24°C for 18 hours in light, 18°C for 6 hours in dark). Three days after planting, the seedlings began to sprout and the plastic covering was removed. The seedlings were then watered every second day, without using fertilizer.

**Seedling Inoculation**

Table 1 outlines the race-by-variety interactions to be sequenced in this project. Each interaction was done with two biological replicates resulting in 14 pots of wheat being infected. Approximately 60, depending on level of seed germination, seven-day old seedlings were inoculated with different rust urediniospores for each pot, specific to variety (Table 1). This was accomplished by adding approximately 100mg of rust spores to 700µL of Bayol oil and spraying this mixture on to the leaves of the plants using a specialized spore inoculator (Geoffrey Harms, Lab Machinist Specialist, U. of Minnesota) (Figure 5) attached to an air compressor at a low pressure (approximately 20 psi).
Figure 5: Custom sprayer attachment. (a) Side view of custom-built, specialized sprayer next to a water soluble capsule used to hold the spore/oil mixture. (b) Front view of the custom-built sprayer next to the same capsule. This view shows the small nozzle. (c) Specialized sprayer with the half capsule inserted as well as the adaptor (gold portion) to hook into an air compressor.
Once sprayed, the seedlings were left to air dry for 30 minutes to allow the oil to dry. The pots were then placed into humid, plastic buckets. These buckets had been prepared before wheat infection by adding approximately half an inch of hot water into the plastic bucket and sealing the bucket with a plastic covering. These were left in the growth chamber over night to allow the humidity to build up to approximately 80-90%. The humidity was measured by using VWR Traceable Hygrometer/Thermometer/Dew Point Monitor. The plastic buckets containing the pots were placed back into the growth chamber, which was set to 18°C for 24 hours in the dark. After 24 hours, the pots were removed from the humidity bins and placed back into the growth chamber at the following conditions: 24°C for 18 hours in light, 18°C for 6 hours in the dark. The plants were monitored for growth and were watered once every second day, without fertilizer.

**Sample Collection**

Leaves from infected plants were collected by cutting a visibly infected leaf at the base. The leaves were weighed and collected in 15mL falcon tubes, with each tube containing 0.2g of leaf tissue (approximately 15-20 leaves) from each of the 14 pots. These leaves were flash frozen using liquid nitrogen and stored at -80°C. The following time points were used for harvesting: 5 days post infection (DPI), 7DPI, 10DPI and 14 DPI. After 14 days, the remainder of the plants were autoclaved and discarded.
RNA Isolation

Isolation of total RNA was performed on 0.2g (15-20) of infected wheat leaves.

Table 1 includes the wheat variety by rust race infections from which RNA was isolated.

Table 1: Rust race and host combinations used in infection time courses. DPI = Days post infection; 4 letter nomenclature system for Rust races developed by Long and Kolmer (1989) (first three letters) and McCallum et al. (2010) (fourth letter); Thatcher is a wheat variety lacking specific leaf rust resistance, the other Lr designations are Thatcher derived varieties with specific leaf rust (Lr) resistance genes bred in.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MFDS x THATCHER 5 DPI</td>
<td>MFDS x THATCHER 7 DPI</td>
</tr>
<tr>
<td>MFDS x Lr3 5DPI</td>
<td>MFDS x Lr24 5DPI</td>
</tr>
<tr>
<td>MFDS x Lr24 7DPI</td>
<td>MBDS x Lr24 7DPI</td>
</tr>
<tr>
<td>MBDS x Lr24 10DPI</td>
<td></td>
</tr>
</tbody>
</table>

The harvesting time points were chosen based on experiments performed by collaborator Xiben Wang (Wang et al., 2013) where percent of fungal mass was roughly determined using PCR based methods on harvested infected plant material at different time points for various race-by-variety interactions, including those indicated in Table 1.

For each leaf sample, a separate mortar and pestle was used to grind the leaves into powder. These were first baked at 300°C for 16 hours in aluminum foil in order to
sterilize the equipment, left to cool to room temperature and were then placed in a -80°C freezer approximately 10 minutes before starting the RNA isolation. This helps to keep the ground plant material from sticking to the mortar or pestle end. The leaf tissue samples (0.2g) were placed in the cold mortar and ground to dust in liquid nitrogen. The frozen ground sample was split in approximately half, placed in to two 1.5mL microcentrifuge tubes and the liquid nitrogen was allowed to evaporate. The kit used requires a maximum of 100mg of plant material.

RNA was isolated from 100mg of plant material using the Qiagen Plant RNA Minispin Kit following the “Purification of total RNA from plant cells and tissues and filamentous fungi” section, which uses the RLC buffer instead of RLT. Buffer RLC contains guanidine hydrochloride. The kit states that buffer RLT typically make the mycelia of filamentous fungi solidify and make it impossible to collect RNA. As suggested by the Plant RNA Minispin protocol, on-column DNA digestion was performed on the RNA using the Qiagen RNase-free DNase set (#79254). The protocol calls for a 15 minute incubation, which I found to not be optimal for these samples as DNA contamination remained. 20 minutes was used instead, as this eliminated DNA contamination. RNA was eluted using 50µL of DEPC-treated water. The eluent was run over the column a second time to increase RNA yield.

The quantity of RNA obtained was determined using a Thermo Scientific NanoDrop 8000 Spectrophotometer. 1.5µL of sample was loaded directly onto the machine after a DEPC-treated water blank and samples were assessed based on their
260/280 and 260/230 profiles on the ND-8000 v 2.2.1 software. The output of the Spectrophotometer includes a measure for the concentration of the sample in ng/µL, which was multiplied according to the volume, in µL, of the total sample to calculate the total ng or mg present in the sample.

Potential genomic DNA presence in the isolated RNA was assessed by carrying out a PCR amplification on the RNA sample, without an initial first strand synthesis step, using primers from Wang et. al. (2013) targeted to putative housekeeping genes specific to wheat and leaf rust. House-keeping genes from both species had to be used due to the nature of the infected plant tissue RNA samples containing both WLR and wheat, thus DNA presence from either had to be accounted for. Wheat specific primers were designed for the putative housekeeping gene encoding puroindolin β (Genbank, AJ302100), with the following sequences: (forward) TGCAAGGATTACGTGATGGA and (reverse) GCTATCTGGCTCAGCTGCTT. The puroindolin β protein is the molecular basis of wheat grain hardiness and texture. Wheat grain hardiness is one of the most important characteristics in determining the end-use of the wheat. The *Puccinia triticina* specific primers were designed for the putative housekeeping gene succinate dehydrogenase (Broad Institute, PTTG_01208.1) with the following sequences: (forward) GATCCTCGCCATAGGAATCA and (reverse) CATCCTATGTGCATGCTGCT. Succinate dehydrogenase is a protein which participates in the electron transport chain and the citric acid cycle. According to Wang et. al. 2013, the expected sizes for each PCR product is 131bp. PCR was performed using AmpliTaq Gold PCR kit from Applied
Biosystems. The master mix and concentrations of reagents are given in Table 2. The PCR conditions were as follows: 95°C for 2 minutes, 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds and 72°C for 1 minute, and then a final extension at 72°C for 10 minutes and a hold at 4°C. The amplified samples were run out on a 1% agarose gel made with 1x TAE buffer and visualized by ethidium bromide staining (0.3 μg/mL, BioShop) using a Geliance 600 Imaging System (Perkin Elmer). Norgen Biotek's Fullranger DNA ladder was used as a size marker.
Table 2: AmpliTaq master mix.

<table>
<thead>
<tr>
<th>AmpliTaq Gold PCR Kit</th>
<th>Final Concentration</th>
<th>x1 reaction (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Water</td>
<td>-</td>
<td>13.375</td>
</tr>
<tr>
<td>PCR Gold Buffer (10X)</td>
<td>1X</td>
<td>2.5</td>
</tr>
<tr>
<td>Magnesium Chloride (25mM)</td>
<td>3 mM</td>
<td>3</td>
</tr>
<tr>
<td>dNTPs (10 mM)</td>
<td>0.2 mM ea</td>
<td>2</td>
</tr>
<tr>
<td>Forward Primer (5 µM)</td>
<td>0.2 µM</td>
<td>1</td>
</tr>
<tr>
<td>Reverse Primer (5 µM)</td>
<td>0.2 µM</td>
<td>1</td>
</tr>
<tr>
<td>AmpGold Taq (5U)</td>
<td>1.25 U</td>
<td>0.125</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>23 µL</td>
</tr>
</tbody>
</table>

The quality of RNA was assessed through the electrophoretic separation of glyoxalated RNA on an agarose gel using 1x BPTE running buffer, as per Sambrook and Russell (2001). An ssRNA Ladder (New England Biolabs) was used as a size marker. This ladder contained bands at the sizes 9000, 7000, 5000, 3000, 2000, 1000 and 500 bases, with the 3000 base fragment having double intensity to serve as a reference. The gel was run at 100volts for 60 minutes.

RNA was isolated with the intention of submitting samples to a centre for RNA sequencing. To complete the sequencing process, the centre required a minimum concentration as well as a normalization of sample concentration. These RNA samples were therefore normalized (where possible) to 250ng/µL in 25µL as requested by the sequencing center (5µg of RNA minimum).
Samples from Clinical Genomics Centre

RNA-seq carried out at Mt. Sinai Hospital's Clinical Genomics Centre (CGC) utilized Illumina technology to create 75bp paired-end reads. Sample libraries were prepared with TruSeq RNA Sample Preparation Kit (Illumina) and these libraries were run on an Illumina HiSeq 2000. Raw data files, in .fastq format, were received from the company and uploaded into the CLC Genomics Workbench 6.5.1 software directly using the "import illumina" option for subsequent alignment to uploaded reference genome data.

Samples from British Columbia Genome Sequencing Centre (BCGSC)

Xiban Wang, a collaborator on the project, had previously submitted samples to the BCGSC for RNA sequencing which complemented those sent by me to the Clinical Genomics Centre. Table 4 shows both, with sample IDs starting with the designation KM being sent by myself to the CGC and those with the designation BR being sent to the BCGSC. RNA-seq carried out at the BCGSC utilized Illumina technology and were pre-aligned to a *P. tritinctina* reference genome and received as .bam files. These files were converted into their .fastq raw form in order to be uploaded into the CLC Genomics Workbench software. To do this, a linux command line software program called Bam2fastq was used. This program was downloaded from http://www.hudsonalpha.org/gsl/software/bam2fastq.php. The prerequisite library zlib
was downloaded from http://www.zlib.net/ and installed. The Bam2fastq program was unzipped and installed using the following command in the linux terminal:

```
tar -xvzf bam2fastq-version-num.tgz.
```

where ‘version-num’ depends on the current version of the program.

The program was operated through the bam2fastq directory, which I navigated to using the “cd” command in the terminal. Within the directory, the structure for the command used is:

```
./bam2fastq [options] <bam file>
```

From that skeleton command, you can build a command specific to your needs using options outlined in the parameters listed upon startup of the software. For paired-end reads, the following command is used:

```
./bam2fastq –o output_read#.fastq input.bam –no-filtered
```

The –o indicates that the title following that option will be the output file name. After the file name is a #, which is needed for paired end reads. The output will either be read1 or read2. The –no-filtered option indicates that all reads will be extracted from the .bam files. These files were then also uploaded into the CLC Genomics Workbench 6.5.1 software directly using the "import illumina" option for subsequent alignment to uploaded reference genome data.
RNA-seq Analysis

A *Puccinia triticina* reference genome (*P. Triticina* 1-1 BBBD Race 1(V2) supercontigs.fasta) and annotation file (*P. Triticina* 1-1 BBBD Race 1(V2) transcripts.gtf) from “*Puccinia* Group Sequencing Project, Broad Institute of Harvard and MIT (http://www.broadinstitute.org/)” were uploaded into the CLC Genomics Workbench v 6.5.1 program using the “import fasta” action. In order to annotate the genome file, the annotation file was merged with the reference genome in CLC using “toolbox-> Classical Sequence Analysis -> General Sequence Analysis->Annotate with GFF/GTF/GVF file...”.

RNA-seq analysis was run on the raw sequencing files, from either CGC or BCGSC, of one data point using “toolbox->Transcriptonomics Analysis->RNA-Seq Analysis->RNA-Seq Analysis”. Two RNA-seq analysis files (the output from the step before) were combined depending on the comparison of interest. The comparisons included determining virulence gene influence on gene expression, gene expression differences based on the response to a resistance gene and identifying expression difference when comparing specific rust race infections. These comparisons are outlined in Figure 9. The files were combined by running an “experiment” within CLC using: “toolbox->Transcriptonomics Analysis->Set-up Experiment”. Statistical analysis was run on the experiment files by selecting the files, one at a time, and running the following: “toolbox->Transcriptonomics Analysis->Statistical Analysis->On Proportions” using Kal’s test in the parameter settings.
These 'experiment' files were exported as an Excel file and uploaded to MS Access. Once loaded into Access, 'experiment' files were compared by using the Query Design function to form queries with specific criteria. RNA-seq data contains a wealth of mineable data. To make it manageable, I had to identify specifically what I was interested in for this project, which would be significant and large expression differences between each different race-by-variety interaction (a specific fungal race on a specific wheat variety). Therefore, my criteria included genes with expression values greater than zero, which would exclude genes which were not expressed in one of the two compared race-by-variety interactions in that 'experiment' file. Further to that, my criteria also included only genes with fold-change expression of greater than 2 or less than –2 with a P-value of <0.05. Although this cut-off was arbitrarily chosen, it limited the data mining to highly significant expression changes between the two compared race-by-variety interactions of each 'experiment' file. A list of genes was generated for each experiment file which adhered to the criteria I stipulated.

**Identifying Secretion Signals**

The next step carried out was to identify if these highly, significantly and differentially expressed genes were potential *P. triticina* effectors. Effectors are proteins which are secreted into the host plant in order to manipulate the surroundings and aid further fungal establishment. The sequences of each listed gene were run through three
online programs to predict whether they are secreted proteins: SoftBerry’s ProtComp v. 9.0, CBS’s SignalP v. 4.1, and CBS’s TargetP v. 1.1.

The first online program used was ProtComp v 9.0, accessed from: http://linux1.softberry.com/berry.phtml. This program predicts protein localization using sequence homology and Neural Networks. The sequences are assigned localization through homology to databases annotated both experimentally and theoretically (Klee and Ellis, 2005). Protcomp 6.0, a past version, was reported to have an 86% correct predictions rate when tested with 200 protein sequences (http://www.softberry.com/berry.phtml?topic=protcompan&group=help&subgroup=proloc). This tool predicts if a peptide sequence is secreted and where it may be secreted also providing a confidence score from 0 to 10. Under the “Protein Location/patterns/epitops” tab, “ProtComp predict the sub-cellular localization for Animal/Fungi” was selected. The consensus sequence of each RNA-seq file was translated into its protein sequence using the ExPASy Translate tool (http://web.expasy.org/translate/) with the “Compact” output format selected. This protein sequence was copied and pasted into the available field on the Protcomp Protein location section for multiple samples simultaneously by inputting the sequences in FASTA format. The primary output of interest is the Integral Prediction line, which combines the algorithms used in the other predictions and gives a confidence score from 0-10. Those outputs which indicate the protein being secreted extracellularly with a confidence score of 7 or more were chosen to be studied further.
Figure 6: Example of an output from ProtComp v 9.0. Important is the Integral Prediction which shows the protein location to be Extracellular (Secreted) with a score that exceeds 7.

SignalP and TargetP are older programs that also have previously demonstrated high accuracy (Menne et al., 2000). Both programs use Neural Networks as well as Hidden Markov Models (HMM) to predict secreted proteins by assessing the sequence at the N-termini of target sequences. These programs are able to discriminate signal peptides from signal anchors in the N-terminus.

SignalP’s Neural Network output includes the position as well as the probability of the residue most likely to belong to a signal peptide (the S-score max), the average probability of all residues examined to belong to a signal peptide (S-score mean), the position and probability of residues most likely to be the first of the N-terminal in the mature protein (C-score max), and the geometric average of the C-score and smoothed
derivative of the S-score (Y-score). The HMM gives the position and score of the residue with the maximum C-score and mean S-score for the entire sequence which was analyzed.

CBS's SignalP v 4.1, accessed from http://www.cbs.dtu.dk/services/SignalP/, was used to predict the secretion signal of the proteins which were determined to be secreted by the ProtComp program. The protein sequences, translated using ExPASy, were pasted into the submission field in FASTA format and the program was run using the default parameters. The output of this program is in the form of C-, S- and Y-scores. Those proteins with C-, S- and Y- scores which pass the program's calculated threshold (approximately 0.5), as depicted in Figure 3, are considered by the program to be secreted. Any non-secretory proteins will be represented by very low scores, typically around 0.1. As explained by the SignalP website, the C-score is the raw cleavage site score and it is the program distinguishing signal peptide cleavage sites from everything else. The S-score is the signal peptide score and it distinguishes the positions within signal peptides from those positions in the mature part of the protein and from proteins without signal peptides. The Y-score is the combined cleavage site score and it is the average of the C-score and the slope of the S-score which gives a better prediction than C-score alone. The program’s output also includes a ‘YES’ under the “signal peptide?” category if it is predicted to be secreted and a “NO” if not. Finally, the predicted cleavage site is also displayed, indicating the position of cleavage between the identified amino acids.
Figure 7: Example of an output from SignalP 4.1.

The third program used was Target P 1.1, accessed from http://www.cbs.dtu.dk/services/TargetP/, which predicts the subcellular location of proteins. TargetP differs from SignalP by predicting either classically secreted proteins (SP), mitochondrial proteins (mTP) or chloroplastic proteins (cTP) through analysis of the N-terminus. The program’s first layer predicts SP, mTP or cTP localization and then the
second layer of the program integrates the localization predictions into a final prediction (Loc). A reliability class (RC) is also assigned to each prediction based on the difference between the highest scoring prediction and the second highest. Again, the translated protein sequences were pasted into the submission window in FASTA format and the default parameters for Non-plant organisms were used. The Loc's predicted localization can either be chloroplast (C), mitochondrion (M), secretory (S) or –, indicating any other location. The program will output an asterisk (*) in that field if the localization cannot be determined. The RC is indicated with a value of 1 through 5, with 1 being the strongest prediction. For these data sets, outputs indicating a strong likelihood of secretion, S1, S2 or S3, were desired as a minimum.

![Image](image.png)

Figure 8: Example of a TargetP 1.1 output.
**Single Nucleotide Polymorphism (SNP) Identification**

Effector proteins are, in addition to being secreted, also identified as usually being approximately 300 amino acids or less (Duplessis *et. al.* 2011). Therefore, proteins which were identified as being secreted in all three programs were then sorted by length, keeping only those which were 300 amino acids or less for further analysis. As indicated in the introduction of this work, due to the rare presence of the secondary host, *P. triticina* does not often undergo sexual recombination to obtain genetic variation. Instead, a mutation or a step-wise series of mutations in the genetic code, known as SNPs, is more likely the cause of genetic variation within races of rust. Thus, the next step in this project was to identify SNPs within the potential effectors identified. The raw sequencing data files of the potential effectors identified to be under 300 amino acids in length were examined in the CLC Genomics Workbench software and those with single nucleotide differences in one race-by-variety interaction verses a second, different interaction, were identified. These comparisons are depicted in Figure 9, which outlines the *in silico* comparisons performed in order to narrow the focus of this investigation from whole RNA-seq data sets to those genes which are highly expressed in one interaction versus another, are also potentially secreted effector proteins and contain mutations. The top box illustrates the comparisons that were used to determine the influence that the different virulence genotypes had on expression while infecting a universal susceptible variety of wheat, Thatcher. BBBD, a WLR race with low virulence, on Thatcher was used as a baseline “control” reaction relative to either MFDS or MBDS on Thatcher. The middle
box depicts *in silico* subtractions where genes specific to MFDS and MBDS infections on Thatcher-\textit{Lr24} were identified. This analysis was completed to identify expression in the two genotypes of interest in response to the resistance gene of interest, \textit{Lr24}, relative to no resistance gene. The bottom box depicts another *in silico* comparison which enables the identification of genes expressed by the newly emerged genotype MFDS on Thatcher-\textit{Lr24} verses that same genotype relative to another resistance gene, in this case, Thatcher-\textit{Lr3}. 
Figure 9: In silico comparisons of WLR races on different wheat varieties and the outcomes expected from those analyses. Wheat varieties Thatcher-\(Lr24\) (labeled as \(Lr24\) only) and Thatcher-\(Lr3\) (labeled as \(Lr3\)) both contain resistance genes, while Thatcher is a universal susceptible variety, containing no resistance gene. WLR race BBBD is a universally weak infector with low virulence seen on even Thatcher. In this experiment, it was used as a baseline comparison.
Finally, the sequences with detected SNPs differing from one data set relative to another were investigated as to their impact on the total protein sequence and structure. Impact on protein sequence was determined by identifying the specific amino acid associated with the SNP and examining if the amino acid was different between the two compared data sets. Changes in these amino acids were identified as either synonymous or nonsynonymous substitutions. Those with nonsynonymous substitutions were checked for alterations in secondary structure of the protein. To do this, translated sequences were examined through the CLC Genomics Workbench software by first loading the sequence using the action: new->sequence->choose protein and save. Next, the software was used to predict protein structure by using the action: toolbox-> classical sequence analysis -> protein analysis -> predict secondary structure. The output of this action depicts the location of alpha-helices and beta sheets within the protein's secondary structure and subsequent comparisons of the output between data sets identified shifts in the locations of these helices and sheets.

Figure 10 outlines the entirety of analyses completed on the RNA-seq data, illustrating how the data was initially focused based on descriptions of effector proteins. The end of the flow chart, Functional characterization, will be the target for future work done with these data sets and the outlined genes of interest based on the analysis.
Figure 10: Flow chart depicting the analyses on the RNA-seq data.
CHAPTER 4: RESULTS

Tissue Growth and Sample Collection

Infected leaf tissue was isolated from multiple race-by-variety interactions, as outlined in Table 1. This tissue was used for RNA isolation and subsequent submission for RNA-seq. The initial obstacle to performing RNA-seq from infected plant tissue is gathering enough fungal material either from early time points of infection or from those infections which are subtle. Two techniques were employed to select for high concentrations of fungal material. First, tissue harvesting included only sites of infection or those areas obviously infected. It was from this that plant and fungal RNAs were isolated. Second, infected tissue was harvested at later days post infection if the infection phenotype was subtle. Wheat leaf rust presents itself as either a resistant or virulent infection type and is scored accordingly. Chlorosis or major flecking of the leaf surface is scored as 0-3, which can be seen in Figure 11 A and B. Wheat leaves with obvious uredinia filled with urediniospores, which are rust or brown coloured and often look or feel powdery, score as a 4-5 as seen in Figure 11 A and C. Wang et. al. (2013) indicated that the amount of fungal biomass in tissue samples increased as infection proceeded. Furthermore, this amount reached its maximum at different times depending on the virulence of the interaction. Based on these results, tissue was harvested at 7 or 10 DPI for infection phenotypes with scores 0-3 and 5 DPI for those infection phenotypes scored as 4-5, as outlined in Table 1.
Figure 11: Infection phenotypes of WLR. A) Taken from Bolton, 2008. Leaves in descending infection type 5 (top) to type 0 (bottom). B) MBDS x ThatcherLr24 7DPI. The yellow flecking on the leaves show a typical 0-3 phenotype, or resistant type infection. C) MFDS x ThatcherLr24 5DPI. The brown coloured spore-filled pustules are typical of a 4-5 phenotype, or virulent type infection.

RNA Isolation

RNA was isolated using the Plant RNA Minispin Kit (QIAGEN) using a DNase I on-column treatment. The RNA was eluted in a small volume to maintain a higher concentration for subsequent library preparation. The concentration of RNA and presence of potential contaminants was determined using a NanoDrop Spectrophotometer. The amount of RNA obtained ranged from 2.9 to 10 mg (Table 3). The absorbance ratio of
260nm and 280nm (A260/280) indicates the purity of RNA (typically 2) with lower ratios (<2) indicate contaminating proteins, residual phenols or other reagents associated with the extraction. A260/280 of RNA samples are shown in Table 3.

**Table 3: NanoDrop Spectrophotometer output** including sample concentration, purity and calculated theoretical recovery.

<table>
<thead>
<tr>
<th>Sample</th>
<th>ng/µL</th>
<th>260/280</th>
<th>Total ng</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFDS x Thatcher 5DPI A</td>
<td>484.7</td>
<td>2.06</td>
<td>14490</td>
</tr>
<tr>
<td>MFDS x Thatcher 5DPI B</td>
<td>495.6</td>
<td>2.04</td>
<td>14868</td>
</tr>
<tr>
<td>MFDS x Thatcher 7DPI A</td>
<td>233.8</td>
<td>2.10</td>
<td>7014</td>
</tr>
<tr>
<td>MFDS x Thatcher 7DPI B</td>
<td>228.4</td>
<td>2.06</td>
<td>6852</td>
</tr>
<tr>
<td>MFDS x Thatcher <em>Lr</em>3 5DPI A</td>
<td>501.3</td>
<td>2.03</td>
<td>15039</td>
</tr>
<tr>
<td>MFDS x Thatcher <em>Lr</em>24 5DPI A</td>
<td>381.9</td>
<td>2.07</td>
<td>11457</td>
</tr>
<tr>
<td>MFDS x Thatcher <em>Lr</em>24 7DPI A</td>
<td>333.1</td>
<td>2.09</td>
<td>9993</td>
</tr>
<tr>
<td>MFDS x Thatcher <em>Lr</em>24 7DPI B</td>
<td>206.6</td>
<td>2.08</td>
<td>6198</td>
</tr>
<tr>
<td>MBDS x Thatcher <em>Lr</em>24 7DPI A</td>
<td>365.0</td>
<td>2.12</td>
<td>10950</td>
</tr>
<tr>
<td>MBDS x Thatcher <em>Lr</em>24 7DPI B</td>
<td>439.2</td>
<td>2.11</td>
<td>13176</td>
</tr>
<tr>
<td>MBDS x Thatcher <em>Lr</em>24 10DPI A</td>
<td>236.1</td>
<td>2.11</td>
<td>7083</td>
</tr>
<tr>
<td>MBDS x Thatcher <em>Lr</em>24 10DPI B</td>
<td>217.7</td>
<td>2.12</td>
<td>6531</td>
</tr>
</tbody>
</table>

The level of intact RNA was determined by visualization following electrophoretic separation on a 1.5% BPTE agarose gel (Figure 12). The presence of an upper large ribosomal subunit band on the gel with an expected intensity 2 fold greater
than the small subunit band (second band) indicated a high level of RNA integrity. The degree of band smearing gives a further indicates of the degree of RNA degradation.

**Figure 12: 1.5% BPTE agarose gel depicting quality of RNA samples.** A) 1µg of RNA was run in each lane along with a marker (M) (ssRNA Ladder, NEB). The gel was visualized on a transilluminator with a 100ms exposure. * indicates sample of insufficient quality. B) Separate electrophoresis of replicates of Sample 12 from A and marker (M). 1µg of RNA was run in the first lane and 0.75µg in the second due to lower sample concentration. Visualization of gel at 150ms.

Each lane of RNA had four distinct bands at ~3500 bp, ~1700 bp, ~1500 bp and ~900 bp. The sample present in the last lane (MBDS x Thatcher Lr24 10DPI, indicated by a *) on the gel was the only sample in which the bands were not distinct but rather smeared. Sample 12, found in the final lane of Figure 12A, was re-isolated and run on a new BPTE gel (Figure 12B). During this attempt, two replicates were isolated. The newly isolated samples were less degraded than the original, though the second sample.
had a low concentration. Therefore it was the first of the two which were used in subsequent analysis.

**RNA-Seq Quality Control**

The quality of RNA was independently assessed at the Mount Sinai Genome Sequencing Center. The output of their analysis, Table 4, includes an RNA Quality Index (RQI). The sequencing centre had previously determined that an RQI of 7 or greater, with 10 being maximum, yield high quality sequencing results in their hands.
Table 4: Quality Control (QC) results for samples submitted for RNA-seq. All samples passed with at least “OK” or a RQI of 7 and were deemed to be of sufficient quality for in-depth sequencing, with "Good" or RQI of 8 or more being preferable.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Sample Name</th>
<th>RQI</th>
<th>QC Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>KM1</td>
<td>MFDS-Thatcher-5DAI-R1</td>
<td>8.1</td>
<td>Good</td>
</tr>
<tr>
<td>KM2</td>
<td>MFDS-Thatcher-5DAI-R2</td>
<td>7</td>
<td>OK</td>
</tr>
<tr>
<td>KM3</td>
<td>MFDS-Thatcher-7DAI-R1</td>
<td>7.3</td>
<td>OK</td>
</tr>
<tr>
<td>KM4</td>
<td>MFDS-Thatcher-7DAI-R2</td>
<td>7.5</td>
<td>OK</td>
</tr>
<tr>
<td>BR5</td>
<td>MFDS-ThatcherLr2a-10DAI-R1</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>BR6</td>
<td>MFDS-ThatcherLr2a-10DAI-R2</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>BR7</td>
<td>MFDS-ThatcherLr2a-14DAI-R1</td>
<td>7.9</td>
<td>OK</td>
</tr>
<tr>
<td>BR8</td>
<td>MFDS-ThatcherLr2a-14DAI-R2</td>
<td>7.9</td>
<td>OK</td>
</tr>
<tr>
<td>KM5</td>
<td>MFDS-ThatcherLr3-5DAI-R1</td>
<td>7.9</td>
<td>OK</td>
</tr>
<tr>
<td>BR10</td>
<td>MFDS-ThatcherLr3-5DAI-R2</td>
<td>7.8</td>
<td>OK</td>
</tr>
<tr>
<td>BR11</td>
<td>MFDS-ThatcherLr3-7DAI-R1</td>
<td>7.2</td>
<td>OK</td>
</tr>
<tr>
<td>BR12</td>
<td>MFDS-ThatcherLr3-7DAI-R2</td>
<td>8.1</td>
<td>Good</td>
</tr>
<tr>
<td>KM6</td>
<td>MFDS-ThatcherLr24-5DAI-R1</td>
<td>8</td>
<td>Good</td>
</tr>
<tr>
<td>BR14</td>
<td>MFDS-ThatcherLr24-5DAI-R2</td>
<td>7.8</td>
<td>OK</td>
</tr>
<tr>
<td>KM7</td>
<td>MFDS-ThatcherLr24-7DAI-R1</td>
<td>7.1</td>
<td>OK</td>
</tr>
<tr>
<td>KM8</td>
<td>MFDS-ThatcherLr24-7DAI-R2</td>
<td>7.6</td>
<td>OK</td>
</tr>
<tr>
<td>KM9</td>
<td>MBDS-ThatcherLr24-7DAI-R1</td>
<td>8.1</td>
<td>Good</td>
</tr>
<tr>
<td>KM10</td>
<td>MBDS-ThatcherLr24-7DAI-R2</td>
<td>8.2</td>
<td>Good</td>
</tr>
<tr>
<td>KM11</td>
<td>MBDS-ThatcherLr24-10DAI-R1</td>
<td>8.1</td>
<td>Good</td>
</tr>
<tr>
<td>KM12</td>
<td>MBDS-ThatcherLr24-10DAI-R2</td>
<td>8.2</td>
<td>Good</td>
</tr>
<tr>
<td>BR21</td>
<td>BBBBD-ThatcherLr24-7DAI-R1</td>
<td>7.5</td>
<td>OK</td>
</tr>
<tr>
<td>BR22</td>
<td>BBBBD-ThatcherLr24-7DAI-R2</td>
<td>7.9</td>
<td>OK</td>
</tr>
<tr>
<td>BR23</td>
<td>BBBBD-ThatcherLr24-10DAI-R1</td>
<td>7.6</td>
<td>OK</td>
</tr>
<tr>
<td>BR24</td>
<td>BBBBD-ThatcherLr24-10DAI-R2</td>
<td>7.7</td>
<td>OK</td>
</tr>
</tbody>
</table>
**Transcriptome Sequencing and Mapping**

Three *Puccinia triticina* races with virulence phenotypes BBBD, MBDS and MFDS were inoculated onto seedlings of Thatcher and Thatcher near-isogenic lines with leaf rust resistance genes *Lr24* and *Lr3*. The transcriptomes of these wheat infections were sequenced using a paired-end Illumina RNA-seq strategy. Looking at the first biological replicate of the samples, the Illumina data returned 478,087,722 reads for the 6 transcriptomes collectively. These reads were mapped to a *Puccinia triticina* reference genome (P. Triticina 1-1BBBD Race 1(V2), Broad Institute) using CLC Genomics Workbench. Mapping statistics are shown in Table 5.

**Table 5: Mapping statistics of paired reads for each sample.** The paired data contains two reads in a pair. Included are specific samples used for further data mining.

<table>
<thead>
<tr>
<th>Name</th>
<th>Total Sequences</th>
<th>Total Sequences Mapped</th>
<th>Percent Mapped (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFDS-Thatcher<em>Lr24</em> 7DPI</td>
<td>88 153 070</td>
<td>10 533 212</td>
<td>11.95</td>
</tr>
<tr>
<td>MBDS-Thatcher<em>Lr24</em> 7DPI</td>
<td>110 094 840</td>
<td>4 359 090</td>
<td>3.96</td>
</tr>
<tr>
<td>MFDS-Thatcher 7DPI</td>
<td>87 560 798</td>
<td>9 413 230</td>
<td>10.75</td>
</tr>
<tr>
<td>MBDS-Thatcher 7DPI</td>
<td>68 101 544</td>
<td>21 446 860</td>
<td>31.49</td>
</tr>
<tr>
<td>BBBD-Thatcher 5DPI</td>
<td>34 287 768</td>
<td>8 982 248</td>
<td>26.2</td>
</tr>
<tr>
<td>MFDS-Thatcher<em>Lr3</em> 7DPI</td>
<td>89 889 692</td>
<td>11 509 108</td>
<td>12.8</td>
</tr>
</tbody>
</table>

The Percent Mapped shows the amount of fungal material within the samples that mapped to the fungal reference genome. Recall that the samples from infected leaf tissue
contain both plant and fungal RNA and therefore a low ‘percent mapped’ indicates a low percentage of fungal RNA in the samples used. The percent mapped values roughly correspond with the susceptibility of the disease interaction and phenotype of infection.

**Normalization of Experiment Data**

Transcripts which aligned to the reference genomic sequences were further analyzed using the CLC Genomics Workbench software package to create ‘Experiment’ data files. These files are comparisons between the data of two samples, reporting information such as fold change and RPKM (reads per kilobase of transcript per million mapped reads) values. In order to compare Experiment files, they were first normalized within CLC.

The normalization of ‘Experiment’ data files was completed within the CLC software and was depicted with box plots. Figure 13A shows an example of expression data for MFDS x ThatcherLr24 and MBDS x ThatcherLr24. Shown is difference in the two plots, red and green. Figure 13B shows the sample plot after normalization using RPKM. The program normalizes by dividing by the total number of reads, which removes most of the variability seen in 13A.
Figure 13: An example of normalization of expression data. A) Box plot of expression data for MFDS x *ThatcherLr24* and MBDS x *ThatcherLr24* show differences. B) Box plot of normalized expression data for MFDS x *ThatcherLr24* and MBDS x *ThatcherLr24* indicate a successful alignment. Normalization plots for all data points found in Appendix 1.

These data sets were assessed for statistical significance in their respective experiment comparisons by performing a Kal’s Z-test on proportions. This significance was represented using volcano plots showing the spread of transcript expression on the x-axis and the p-value along the y-axis. Figure 14 shows a volcano plot for data relating to MFDS x *Thatcher Lr24* vs. MBDS x *Thatcher Lr24.*
Figure 14: A volcano plot of expression data MFDS x Thatcher Lr24 vs MBDS x Thatcher Lr24. This plot shows the spread of transcript expression relative to the p-value. The higher the data point, the higher the significance of the data point. The further left or right from the middle, the higher the fold change. Volcano plots for all expression data are shown in Appendix 2.

Expression Analysis and Identification of Candidate Effector Genes

The lists of genes expressed in the ‘experiment’ files for each comparison were then examined. First, I excluded any transcripts with no expression in both of the data sets. These were removed as this specific project focused on expressed effectors potentially altering virulence. Second, the data sets were narrowed further to those
transcripts with a greater than two-fold change in expression in each experiment pairing
(Kal’s Z-test p=0.05). Even though any difference may be interesting and important to
virulence variations, the scope of this project required the focus to shift to largely up or
down regulated differences.

Further criteria for candidate selection were based upon the features of previously
characterized effectors. The list of potential candidates was screened for those transcripts
encoding proteins with secretion signals using SoftBerry’s ProtComp v. 9.0, CBS’s
SignalP v. 4.1, and CBS’s TargetP v. 1.1. This resulted in a refined list of over 166
expressed transcripts with predicted secreted proteins. The translated protein sequences of
these transcripts were examined and those with more than 300 amino acids were excluded
from further analysis. The remaining 97 candidate effector-encoding genes with a two-
fold or higher change in transcript level are presented in Table 6 (green column).
**Table 6: Potential effectors and those with SNPs resulting in changes to the amino acid sequences.** 97 candidate effectors with a 2-fold or higher change in transcript level were identified (green column). 12 of these candidate effector genes were recognized to contain SNPs which altered the amino acid sequence (yellow column).

*Protein sequences that passed a minimum threshold of confidence in calling the secretion signal for two of the programs used: Score 7 or higher in ProtComp, and S1-S3 in TargetP.

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Expression &gt;0, Fold Change &gt;2 or &lt;-2, P-Value &lt;0.05</th>
<th>Secreted</th>
<th>Secreted with High Score*</th>
<th>SNP</th>
<th>Amino Acid Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>MBDS on Thatcher ( Lr24 ) – MBDS on Thatcher</td>
<td>161</td>
<td>5</td>
<td>2</td>
<td>Low expression</td>
<td>--</td>
</tr>
<tr>
<td>MFDS on Thatcher ( Lr24 ) – MFDS on Thatcher</td>
<td>548</td>
<td>13</td>
<td>8</td>
<td>Low expression</td>
<td>--</td>
</tr>
<tr>
<td>MFDS on Thatcher ( Lr24 ) – MFDS on Thatcher ( Lr3 )</td>
<td>415</td>
<td>16</td>
<td>11</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MFDS on Thatcher ( Lr24 ) vs MBDS on Thatcher ( Lr24 )</td>
<td>677</td>
<td>14</td>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MFDS on Thatcher vs MBDS on Thatcher</td>
<td>994</td>
<td>20</td>
<td>10</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>MBDS on Thatcher vs BBBD on Thatcher</td>
<td>3451</td>
<td>50</td>
<td>28</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>MBDS on Thatcher ( Lr24 ) vs MBDS on Thatcher</td>
<td>1434</td>
<td>34</td>
<td>18</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>MFDS on Thatcher ( Lr24 ) vs MFDS on Thatcher ( Lr3 )</td>
<td>428</td>
<td>12</td>
<td>8</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>MFDS on Thatcher ( Lr24 ) vs MFDS on Thatcher</td>
<td>27</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Investigation of Amino Acid Changes

The DNA sequences of the candidate effectors were compared within the pairings with the purpose of identifying single nucleotide polymorphisms (SNPs), specifically those SNPs which altered the resulting amino acid sequence of the ORF in a way which impacted the structure of the predicted protein. 12 of the candidate effector genes were recognized to contain SNPs which altered the amino acid sequence of the protein (Table 6, yellow column). In order to change the sequence of the amino acid, the SNP must be a nonsynonymous substitution of the base. Table 7 outlines the change of the amino acid for each gene, as well as the location in the gene sequence that the SNP occurs. Table 7 also shows that some of the 12 genes are the same in different infection profiles, resulting in 7 unique genes found across the profiles.
Table 7: Nonsynonymous SNPs present in candidate effector genes. Listed is the specific amino acid changes and their location in the protein sequence for each of the 12 putative effectors with amino acid changes outlined in Table 6. *protein no longer in frame due to insertion, ? = deletion of amino acid

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Gene</th>
<th>Location of Change (aa)</th>
<th>Size of Protein (aa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFDS on Thatcher Lr24 – MFDS on Thatcher Lr3</td>
<td>PTTG_28409</td>
<td>T23A</td>
<td>72</td>
</tr>
<tr>
<td>MFDS on Thatcher Lr24 vs MBDS on Thatcher Lr24</td>
<td>PTTG_27691</td>
<td>Q252K</td>
<td>135</td>
</tr>
<tr>
<td>MFDS on Thatcher vs MBDS on Thatcher</td>
<td>PTTG_28503</td>
<td>R28Q</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>PTTG_30866</td>
<td>A26T</td>
<td>255</td>
</tr>
<tr>
<td>MBDS on Thatcher vs BBBD on Thatcher</td>
<td>PTTG_30866</td>
<td>A26T</td>
<td>255</td>
</tr>
<tr>
<td>MBDS on Thatcher Lr24 vs MBDS on Thatcher</td>
<td>PTTG_26225</td>
<td>V33I</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>PTTG_27148</td>
<td>R129L</td>
<td>234</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E133?</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTTG_27221</td>
<td>G88A</td>
<td>115</td>
</tr>
<tr>
<td></td>
<td>PTTG_27691</td>
<td>S78G H112R</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Insertion 337/8 bp*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTTG_28503</td>
<td>R28Q</td>
<td>176</td>
</tr>
<tr>
<td></td>
<td>PTTG_30866</td>
<td>A26T R164G</td>
<td>255</td>
</tr>
<tr>
<td>MFDS on Thatcher Lr24 vs MFDS on Thatcher Lr3</td>
<td>PTTG_30866</td>
<td>A26T R164G</td>
<td>255</td>
</tr>
</tbody>
</table>
The WLR genome is in the process of being annotated, but as it stands, most of the identified potential effector genes are indicated as coding for hypothetical proteins. All of the 12 genes of interest were hypothetical. In order to understand if these genes are potentially novel or if they have identified orthologs in other species, searches using BLASTx were completed. The first search was against the genome of the well-studied, biotrophic model pathogen *Ustilago maydis* and a second search open to all organisms and the highest scoring hit was recorded. Potential orthologs to these 12 genes were all hypothetical proteins, indicating the potential for these genes to be novel effectors (Table 8). Most BLASTx results hit other related fungi, with one gene, PTTG_28409, hitting a plant interacting bacterium *Sinorhizobium meliloti*, though none of those hits had a significant E-value.
Table 8: Potential orthologs to the putative effector proteins. All potential orthologs were hypothetical in both *U. maydis* as well as the closest hit on any organism.

NS – Non significant E-value

PTTG – *Puccinia triticina* (wheat leaf rust)

PGTG – *Puccinia graminis f.sp. tritici* (wheat stem rust)

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Gene</th>
<th>BLASTx</th>
<th>E-value</th>
<th>BLASTx <em>U. maydis</em></th>
<th>E-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFDSxThat – MBDSxThat</td>
<td>PTTG_28503</td>
<td>PGTG_00346 – hyp protein</td>
<td>4E-94</td>
<td>UMAG_01670 – hyp protein</td>
<td>1E-32</td>
</tr>
<tr>
<td></td>
<td>PTTG_30866</td>
<td>PGTG_21151 – hyp protein</td>
<td>2E-09</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>MBDSxThat vs BBBDxThat</td>
<td>PTTG_30866</td>
<td>PGTG_21151 – hyp protein</td>
<td>2E-09</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>MFDSxLr24 vs MBDSxLr24</td>
<td>PTTG_27691</td>
<td>PGTG_02680 – hyp protein</td>
<td>7E-32</td>
<td>UMAG_03238 – hyp protein</td>
<td>0.002</td>
</tr>
<tr>
<td>MBDSxLr24 vs MBDSxThat</td>
<td>PTTG_26225</td>
<td>NS</td>
<td></td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTTG_27148</td>
<td>PGTG_02178 – hyp protein</td>
<td>2E-30</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTTG_27221</td>
<td>PGTG_07260 – hyp protein</td>
<td>5E-07</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTTG_27691</td>
<td>PGTG_02680 – hyp protein</td>
<td>0.032</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>PTTG_28503</td>
<td>PGTG_00346 – hyp protein</td>
<td>4E-94</td>
<td>UMAG_01670 – hyp protein</td>
<td>1E-32</td>
</tr>
<tr>
<td></td>
<td>PTTG_30866</td>
<td>PGTG_21151 – hyp protein</td>
<td>2E-09</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>MFDSxLr24 vs MFDSxLr3</td>
<td>PTTG_30866</td>
<td>PGTG_21151 – hyp protein</td>
<td>2E-09</td>
<td>NS</td>
<td></td>
</tr>
<tr>
<td>MFDS on Thatcher <em>Lr24</em> – MFDS on Thatcher <em>Lr3</em></td>
<td>PTTG_28409</td>
<td>NS</td>
<td></td>
<td>NS</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 5: DISCUSSION

Races of wheat leaf rust are constantly changing and overcoming resistances bred into the currently used wheat varieties. In Ontario, the wheat variety Vienna with the resistance gene\textit{Lr24} was overcome by a race of rust with genotype MFDS. It was hypothesized that the previously prevalent rust race MBDS, avirulent on \textit{Lr24}, underwent a change in its genome, creating the new race MFDS. This change was considered to be a point mutation(s) or SNP(s) rather than the result of recombination since the alternate, sexual host, \textit{T. speciosissimum}, is rarely found in North America's wheat growing areas. To identify changes which have an impact on gene expression during pathogenesis of the wheat host, transcriptomic changes among races of wheat leaf rust during infection were analyzed for sequence variation.

\textbf{Transcriptome Sequencing and Mapping}

To obtain a list of genes that are differentially expressed in one race-by-variety interaction versus another, infected leaf tissue samples were collected from seedlings. The maturity of the leaf selected for harvest was determined in previous experiments by collaborator, Xiben Wang (2013), which estimated the relative levels of fungal to wheat biomass in samples at time points in various infection types and wheat cultivars. He determined that the amount of fungal biomass in the samples increased differentially with
rust infections of different virulence levels. The time points selected for RNA isolation were those DPI at which the fungal biomass fraction was at its maximum.

RNA sequencing using Illumina technology was performed to provide transcriptomes of different race-by-variety interactions. These transcriptomes were used to identify expression profile differences and candidate virulence genes. The output paired sequences were mapped to the *P. triticina* reference genome as laid out in Table 5. The Percent Mapped shows the amount of fungal material mapped to the fungal reference genome. The generally low percentage mapped is expected for all samples as the source is material from infected plant samples from early infection time points. Infected plant samples yield lower fungal fraction in isolated RNA samples than plant fraction. Interestingly, the percent mapped values roughly correspond with the susceptibility of the disease interaction, showing lower percentages for those samples from infections on resistant varieties of wheat, for example, showcasing the resistance Thatcher*Lr24* has against race MBDS with the low fungal presence in the sample. The resistant plants prevent fungal proliferation and thus those samples will have a lower fraction of fungal material present.

The genome size of *P. triticina* is 135 Mb (broad institute genome statistics) with 14,878 predicted protein-coding genes. The size of the proteome is somewhat smaller than close relatives *Puccinia graminis* f. sp. *tritici* (17,773) and *Melampsora larici-populina* (16,399) (Duplessis *et al.*, 2011) but considerably larger than either *Ustilago maydis* or *Sporisorium reilianum*, two nonobligate pathogenic biotrophs with proteomes
of approximately 6500 (Kamper et al., 2006; Schirawski et al., 2010). However, the proteome itself may not be the most useful when trying to identify potential effectors. More specifically, the secretome, or those proteins which are secreted, may be of more use for study as effectors are classified as being secreted into the host plant. It has been demonstrated that secreted peptides, or those proteins found within the secretome, act as avirulence effectors in other pathogenic fungi (Dodds et al., 2004, Nirmala et al., 2011). It has recently been found that the secretome of different races of rust differ slightly. Bruce et al. (2014) identified 76 predicted secreted proteins which differed between 6 different races of rust. This study also identified genes of secreted proteins which had alterations in their amino acids, corresponding to putative effectors which are potentially recognized by 11 different Lr genes (Bruce et al., 2014). The identification of novel proteins with alterations in their genetic material resulting in different or differently expressed proteins mimics the purpose of this thesis. After sequencing of the transcriptome through RNA-seq, expression analysis of the data, as well as identifying genes with secreted proteins, was performed.

**Expression Analysis and Identification of Candidate Effector Genes**

This study generated data in the form of experiment files which were normalized to enable cross comparison, and subsequently used to identify potential effectors differing between infection profiles. Using the web-based programs ProtComp, TargetP and
SignalP, 166 proteins were found to be potentially secreted into the host. To be conservative, a threshold of score 7 on ProtComp and S1-S3 in TargetP was used, resulting in 97 proteins predicted to be secreted. From this pool, it was determined that 19 proteins contain at least one SNP when compared between experiment files and relative to the reference. Bruce et al. 2014 identified 532 predicted secreted proteins from 222,571 contigs of six *P. triticina* races. Of these, 456 were found in all six races studied, leaving 76 secreted proteins found differentially in the six transcriptomes studied. Further, the study identified 15 genes which contained amino acid changes (Bruce et al., 2014). The 19 proteins containing SNPs were further analyzed, identifying those with nonsynonymous substitutions in the DNA sequence, and thus altering the amino acid sequence and potentially the secondary structure of the resulting protein. These nonsynonymous coding SNPs may have the largest impact on phenotype variation. There has been numerous studies in humans regarding predicting phenotypic changes due to found nonsynonymous SNPs in gene regions, as explained in Ng and Henikoff, 2004. SNPs can be associated with Alzheimer's disease and systemic lupus pathogenicity in humans (Karch and Goate, 2015; Bentham et al., 2015). Though these studies do not include SNPs altering effector proteins, a study within the oomycete *Plasmopara halstedii*, causal agent of downy mildew on sunflower, shows SNPs altering effector proteins which results in different pathotypes (Gascuel et al., 2016). These studies aid in setting a precedent that alterations due to SNPs may alter pathogenesis or pathotype of a disease. This is important in this study of WLR as it may be a SNP or gene mutation
which is the reason for the alteration in pathogenesis as sexual reproduction of the fungus is typically absent in North American wheat growing areas.

A gene transcript of particular interest is PTTG_27691, which was found in both MFDS on Thatcher\textit{Lr24} vs MBDS on Thatcher\textit{Lr24} and in MBDS on Thatcher\textit{Lr24} vs MBDS on Thatcher. In the two different interactions, the gene is seen to be variable. There was a single SNP altering the protein at the 252\textsuperscript{nd} amino acid in MBDS on Thatcher\textit{Lr24} relative to MFDS on Thatcher\textit{Lr24}. However, two SNPs were detected in the MBDS on Thatcher iteration, when compared to MBDS on Thatcher\textit{Lr24}, altering the 78\textsuperscript{th} and 112\textsuperscript{th} amino acids. There was also an insertion between the 337\textsuperscript{th} and 338\textsuperscript{th} basepairs in the DNA sequence, resulting in a frameshift mutation, altering the amino acid sequence of the remainder of the protein. The identification of this insertion was supported by a multitude of overlapping RNA-seq fragments. To reiterate, the second instance (as per the list in Table 7) of this gene had two altered amino acids and an insertion disrupting the protein all before the first iteration’s altered amino acid. It would be very insightful to perhaps sequence this gene again in both the MBDS on Thatcher\textit{Lr24} and MBDS on Thatcher for comparison and confirmation of this gene variability. Further, it would be interesting to amplify this gene in a multitude of backgrounds to get a full indication of the potential level of variability of the gene.
In Silico Protein Analysis

The nonsynonymous substitutions were checked for change in charge in side chain/protein surface. Changes in the charge of the protein surface affects how that protein interacts with other proteins and its environment. The charge of the amino acid may also influence how the protein folds, which may also have a direct impact on how the protein interacts with other molecules or its environment. It is well known that alterations protein folding will influence how the molecular or biochemical functions of the proteins are carried out (Buljan et al., 2013). This data contains SNPs such as that seen in the protein of PTTG_27691, which changes from Glutamine, a neutral polar amino acid to Lysine, a negatively charged amino acid. This change in charge can have direct implications on how that section of the protein may fold on itself or/and how it interacts with other proteins. Similarly, the protein of PTTG_28503 was shown to change in the opposite way; from Arginine, a positively charged amino acid, to Glutamine, the polar neutral amino acid. Along with change in charge of the amino acid, the change in size of the amino acid may also impact how that protein interacts with its environment. A change from a small side chain to a large and bulky chain can potentially influence how the protein fits into receptors, which is particularly important in effector protein recognition by the host. The protein of PTTG_28503 contains the change Arginine to Glutamine. Arginine has a larger side chain in general, as well as is less linear than Glutamine. The end of the Arginine chain is bulky and physically takes up more space than the slimmer Glutamine. The change in size of this particular section of the protein
may also influence how the protein folds, as well as influencing how it directly interacts with other proteins.

After identifying changes, these proteins were submitted and searched using BLAST to identify if potential orthologs exist in similar species. The 12 potential effectors identified in each comparison were indicated as coding for hypothetical proteins within *P. triticina*. In order to identify if these putative effectors have potential orthologs in related species, BLASTx searches were completed on the potential effector list, as seen in Table 8. When using BLAST specifically against the biotrophic model pathogen *Ustilago maydis*, potential orthologs were identified for only two WLR proteins, PTTG_28503 is a potential ortholog to UMAG_01670 and PTTG_27691 is a potential ortholog of UMAG_03238. Interestingly, the allele for PTTG_27691 expressed in MBDS grown on Thatcher*Lr24* did not align to the same *U. maydis* protein. Further, the resulting allele was not similar to any protein within *U. maydis*. This result supports the contention that the changes in sequences identified here have the potential to alter the protein sequence in a manner which could change the function or ability of the protein to interact with another protein. Both of the *U. maydis* genes hit were hypothetical proteins and although they contain domains with known functions, neither have been characterized within *U. maydis*. This is fairly significant as *U. maydis* is a well characterized biotrophic fungus, similar to WLR. It could be reasoned that, even with the Smut and Rust species being evolutionarily separate, the similarity in lifestyle could indicate that the proteins involved in directing the fungal interaction with the host could also be similar.
The 12 candidate effectors were also searched using BLASTx without an organism restriction, allowing for any closely related species to provide a hit. This resulted in almost every gene hitting to the related wheat stem rust, *Puccinia graminis* f. sp. *tritici*. All of these stem rust genes translated into hypothetical proteins. This could mean that the proteins identified are novel within WLR, perhaps further indicative of effector proteins, rapidly evolving to overcome resistances in wheat cultivars. It is the opinion of some that effectors are evolutionarily diverse and present as novel, uncharacterized proteins in order to provide the opportunity for undetected invasion of the plant host (Saunders *et al.*, 2011). However, unlike the well-studied *U. maydis* fungi, *P. graminis* is not as well characterized partly due to its intractable nature in a laboratory setting. Thus, the hypothetical definition of the *P. graminis* potential orthologs may be due to incomplete information on the genome and not due to their novelty.

**Future Directions**

There are many avenues of research that this thesis opens for future studies, be it further mining into the RNA-seq database created for this project, or functional characterization of the potential effectors or their orthologs identified in this study. A component of functional characterization of PTTG_27691 should include screening for sequence variation in this gene among other wheat leaf rust races at the DNA level and at the transcript level.
This study identified 7 potential *P. triticina* effector protein genes. Two of these effectors have potential orthologs in *U. maydis* (Table 8). To confirm these, it would be useful to conduct a reciprocal BLAST using *U. maydis*. Also, identifying transit peptide sequences within the *U. maydis* putative orthologs would indicate if the proteins of these genes were capable of being secreted and therefore, potentially being effector proteins themselves. However, effector proteins are often novel to the system, and thus if the *U. maydis* putative orthologs does not present itself as a potential effector, it should not detract from the original analysis in *P. triticina*. However, any sequence conservation may indicate a conservation of function of the specific proteins.

Having putative orthologs in *U. maydis* opens avenues of research unobtainable through the *P. triticina* system, which is not tractable in a laboratory setting. Using the putative orthologs, an expression analysis could be conducted using a cell type panel with the results hopefully showing upregulation *in planta* in relation to the other cell types. This result would indicate that the genes are turned on only while infecting the plant and not during regular growth or maintenance of the fungus, and thus is needed for infection or controlling the infection environment. The genes could also be deleted and over expressed in a *U. maydis* background to assess the impact they may have on virulence. The deletion could be complemented by the *P. triticina* gene directly to see if it is capable of functioning within *U. maydis*.

This thesis includes only a small fraction of information that can be gleaned from further mining of the RNA-seq data. There are many other differences that can be
potentially identified when looking at the different infection profiles. This can include subtle shifts in expression rather than a greater than 2 fold change as was described in this thesis as well as searching for a presence/absence of expression profiles in various race-by-variety infection interactions. Further, the RNA-seq database created for this project can be used to compare with other databases created, including *Puccinia graminis*, stem rust. These comparisons between RNA-seq databases may identify further conserved effectors between, for example, different rust species.

This new database of transcriptomic information could also be used to help annotate the *P. triticina* genome. This could be started by identifying potential orthologs in other closely related organisms which may indicate function of protein for characterization. Since *P. triticina* is not amendable to genetic manipulation in a laboratory setting, cloning genes into a species such as *U. maydis* for subsequent analysis may be possible to overcome this specific obstacle. Further, due to *U. maydis* being a model organism for biotrophic fungi, it may be possible to identify orthologs in this genome and perform expression and functional analyses to infer the potential role of the expressed genes in *P. triticina*.

Obtaining an RNA-seq database allows for many analyses and data-mining experiments, which can elucidate an extraordinary amount of information on the organism sequenced and how it compares to related species. This thesis was only able to glean a small portion of information from the database, but this huge database is now
available for future study of the *P. triticina* infection process and its variation among various race by host interactions.
CHAPTER 6: CONCLUSIONS

Fungal diseases, such as biotrophic rusts, have had a negative economic impact on cereal crop production. *Puccinia triticina*, or wheat leaf rust, is high on the list of damaging rusts. To combat this, wheat cultivars are bred with resistances to current races of rust infecting agricultural fields. These resistances last only a short time until new races of rust overcome them. In Ontario wheat fields, a recent emergence of a WLR race with a new virulence type offered an opportunity to study transcriptome changes in the emerged race relative to a possible ancestor race.

Various race-by-variety interactions were designed to detect variation in candidate virulence genes using RNA-seq technology. The purpose of this study was to identify these candidate virulence genes and provide insight into WLR virulence evolution. A list of genes with statistically significant change in expression between race-by-variety comparisons was generated. From this list, predicted secreted proteins were identified using three web based programs: SignalP, TargetP and ProtComp. 97 candidate WLR virulence genes were identified in all comparisons, 7 of which contained at least one nonsynonymous SNP. These SNPs altered the amino acid sequence and secondary structure of the resulting proteins. One gene, PTTG_27691, was discovered to have multiple sequence variants in different interactions, at least within the context of this investigation. BLASTx searches reported potential orthologs to these 7 proteins within the related rust species *Puccinia graminis* f. sp. *tritici*, wheat stem rust, as well as some having potential orthologs within the model organism for biotrophic fungi, *Ustilago*.
maydis, common smut of corn. All potential orthologs were identified as hypothetical in the other organisms.

The potential orthologs identified within *U. maydis* opens the door to functional analyses of these genes. The fungus *P. triticina* is not tractable in a laboratory setting, however *U. maydis* is. Expression analysis could be conducted using these potential orthologs, as well as gene manipulation to assess the impact on virulence of these proteins. Further study into the gene PTTG_27691 should include screening for sequence variation to confirm the variation identified in this study.

The great wealth of data generated within this project enables many avenues of research in the future. This data can be used to perform other searches, similar to that conducted in this project, but with different criteria, such as presence or absence of expression profiles when comparing expression profiles. Further, this data can be used, in the future, to potentially assist in annotating the *P. triticina* genome. As mentioned, this also may be aided through functional analysis and characterization using *U. maydis* orthologs.
REFERENCES


Ashfield T, Keen NT, Buzzell RI, Innes RW. Soybean resistance genes specific for different Pseudomonas syringae avirulence genes are allelic, or closely linked, at the RPG1 locus. Genetics. 1995;141:1597-604.


Dénes TE, Molnár I, Rákosi-Tican E. New insights into the interaction between cultivated potato and *Phytophthora infestans*. Studio Bio UBB. 2015;1:165-175.


Gascuel Q, Bordat A, Sallet E, Pouilly N, Carrere S, Roux F, Vincourt P, Godiard L. Effector polymorphisms of the sunflower downy mildew pathogen *Plasmopara halstedii*
and their use to identify pathotypes from field isolates. PLOS one. 2016. http://dx.doi.org/10.1371/journal.pone.0148513.


Ng PC, Henikoff S. Predicting the effects of amino acid substitutions on protein function. Genomics and Human Gen. 2004;7:61-80.


APPENDICES
Appendix 2: Volcano plots representing expression data. These plots depict the spread of transcript expression relative to the p-value. The higher the data point, the higher the significance of the data point. The further left or right from the middle, the higher the fold change. A) MBDS x ThatcherLr24 and MBDS x Thatcher. B) MBDS x Thatcher and BBBD x Thatcher. C) MFDS x ThatcherLr24 and MFDS x ThatcherLr3. D) MFDS x ThatcherLr24 and MFDS x Thatcher. E) MFDS x Thatcher and BBBD x Thatcher. F) MFDS x Thatcher and MBDS x Thatcher.