ESTABLISHMENT OF BIOTROPHY BY THE MAIZE ANTHRACNOSE PATHOGEN COLLETOTRICHUM GRAMINICOLA: USE OF BIOINFORMATICS AND TRANSCRIPTOMICS TO ADDRESS THE POTENTIAL ROLES OF SECRETION, STRESS RESPONSE, AND SECRETED PROTEINS

Ester Alvarenga Santos Buiate
University of Kentucky, esterbuiate@uky.edu

Recommended Citation
https://uknowledge.uky.edu/plantpath_etds/17
STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained needed written permission statement(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine) which will be submitted to UKnowledge as Additional File.

I hereby grant to The University of Kentucky and its agents the irrevocable, non-exclusive, and royalty-free license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless an embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s thesis including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Ester Alvarenga Santos Buiate, Student

Dr. Lisa J. Vaillancourt, Major Professor

Dr. Lisa J. Vaillancourt, Director of Graduate Studies
ESTABLISHMENT OF BIOTROPHY BY THE MAIZE ANTHRACNOSE PATHOGEN COLLETOTRICHUM GRAMINICOLA: USE OF BIOINFORMATICS AND TRANSCRIPTOMICS TO ADDRESS THE POTENTIAL ROLES OF SECRETION, STRESS RESPONSE, AND SECRETED PROTEINS

____________________________________

DISSERTATION

____________________________________

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the College of Agriculture, Food and Environment at the University of Kentucky

By

Ester Alvarenga Santos Buiate

Lexington, Kentucky

Director: Dr. Lisa J. Vaillancourt, Professor of Plant Pathology

Lexington, Kentucky 2015

Copyright © Ester Alvarenga Santos Buiate 2015
ABSTRACT OF DISSERTATION

ESTABLISHMENT OF BIOTROPY BY THE MAIZE ANTHRACNOSE PATHOGEN COLLETOTRICHUM GRAMINICOLA: USE OF BIOINFORMATICS AND TRANSCRIPTOMICS TO ADDRESS THE POTENTIAL ROLES OF SECRETION, STRESS RESPONSE, AND SECRETED PROTEINS

Colletotrichum graminicola is a hemibiotrophic pathogen of maize that causes anthracnose leaf and stalk rot diseases. The pathogen penetrates the host and initially establishes an intracellular biotrophic infection, in which the hyphae are separated from the living host cell by a membrane that is elaborated by the host, apparently in response to pathogen signals. A nonpathogenic mutant (MT) of C. graminicola was generated that germinates and penetrates the host normally, but is incapable of establishing a normal biotrophic infection. The mutated gene is Cpr1, conserved in eukaryotes and predicted to encode a component of the signal peptidase complex. How can we explain why the MT is normal in culture and during early stages of pathogenicity, but is deficient specifically in the ability to establish biotrophy? To address this, first I characterized the insertion in the 3' UTR of the MT strain in detail, something that had not been done before. The wild-type (WT) transcript did not differ from predictions, but the MT produced several aberrant transcript species, including truncated and non-spliced transcripts, and the normal one. Aberrant splicing of MT cpr1 was observed both in RNAseq transcriptome data and reverse-transcription polymerase chain reaction (RT-PCR), under different growth conditions and in planta. I also conducted a bioinformatic analysis of other conserved components of the secretory pathway in the MT and WT in planta. One explanation for nonpathogenicity of the MT is that it cannot cope with an increase in secretory activity during infection, and fails to produce necessary pathogenicity factors. With the transcriptome data, I was able to identify effector proteins that were expressed in the WT but not in the MT. Another possible explanation for the MT phenotype is that the MT can't adapt to stress imposed by the plant. I developed a growth assay to characterize the effect of chemical stressors in vitro. The MT was more sensitive to most stressors, when compared to the WT. The transcriptome data indicates that the genes involved in different stress pathways are expressed in planta in both WT and MT, although very few genes are differentially expressed across the different growth stages.

KEYWORDS: corn anthracnose, secretion pathway, fungal stress pathway, fungal effectors, bioinformatics.

Copyright © Ester Alvarenga Santos Buiate 2015
ESTABLISHMENT OF BIOTROPHY BY THE MAIZE ANTHRACNOSE PATHOGEN COLLETOTRICHIUM GRAMINICOLA: USE OF BIOINFORMATICS AND TRANSCRIPTOMICS TO ADDRESS THE POTENTIAL ROLES OF SECRETION, STRESS RESPONSE, AND SECRETED PROTEINS

by

Ester Alvarenga Santos Buiate

Lisa J. Vaillancourt
Director of Dissertation

Lisa J. Vaillancourt
Director of Graduate Studies

March 23, 2015
Date
Aos meus pais, que foram meus primeiros professores com seu exemplo de força e caráter. O constante apoio e amor incondicional de vocês se refletem nas minhas conquistas. Obrigada!

Für Hermann, der mir Stärke gab als ich selbst keine hatte.
Acknowledgements

I’ve known Dr. Lisa Vaillancourt for almost nine years, and she has not only been a great mentor and a colleague, but a great friend. Her patience, concern and dedication to her work and her students, and her faith in me are what motivated me during those years. I am forever grateful for all that you taught me. I would like to thank my dissertation committee, Dr. Mark Farman, Dr. Peter Nagy and Dr. Arthur Hunt, for their academic support and insightful advice over the years. In addition, my gratitude is extended to the faculty members of the department, from whom I learned much through classes and our conversations.

I am extremely grateful to Etta Nuckles and Doug Brown not only for their great work, essential for this dissertation to be completed, but for their encouragement and friendship. To all the staff of the Department, especially (my maid of honor) Alicia Landon and Shirley Harris, many thanks for always been a tremendous help and for the laughs when needed.

During these years at Vaillancourt Lab, I was extremely lucky to have had incredible PhD students as my colleagues. I would like to thank Dr. Sladana Bec, you taught me so many things in the lab, as well as to live life in a sweeter way. Your friendship has been invaluable inside and outside the lab. Dr. Maria Torres, we’ve shared benches, too many corn sheaths, a few worries and incredibly funny moments. I am grateful to have met such an amazing scientist and a life-long friend. Many thanks to Katia Xavier, for always providing the much-needed coffee, encouragement and for being a tremendous help whenever I needed. Your friendship made the completion of this dissertation easier. To all the other past and current members of the Vaillancourt Lab whom I’ve had the pleasure to work with, thank you for your help, conversations and support. Finally, I would like to thank the friends I made over the years at UK, most now scattered around the world. Your support, near or far, was essential.
Table of contents

Acknowledgements ........................................................................................................ iii
Table of contents ............................................................................................................... iv
List of Figures ..................................................................................................................... viii
List of Files ......................................................................................................................... xi

Chapter 1 – Colletotrichum graminicola, the causal agent of maize anthracnose
........................................................................................................................................ 1
1.1 Overview ......................................................................................................................... 1
1.2 The economic impact of maize anthracnose ................................................................. 1
1.3 Disease cycle of anthracnose ....................................................................................... 3
1.4 Hemibiotrophy and its role in the disease cycle in Colletotrichum ...................... 4
1.5 Resistance to maize anthracnose ............................................................................... 7
1.6 The Cpr1 mutant of C. graminicola ........................................................................ 8
1.7 What is the role of Cpr1 in C. graminicola pathogenicity? ................................ 10
1.8 Hypotheses and objectives of this dissertation ....................................................... 11

Chapter 2 – Transcriptional analysis of Cpr1, and other components of the
signal peptidase complex and secretory pathway, in wild type and mutant
strains of Colletotrichum graminicola ........................................................................... 18
2.1 Overview ....................................................................................................................... 18
2.2 Introduction .................................................................................................................. 19
2.3 Material and Methods ............................................................................................... 24
   2.3.1 Defining the putative secretory pathway in C. graminicola ....................... 24
   2.3.2 Sequence analysis of signal peptidase complex proteins ......................... 24
   2.3.3 Fungal strains ....................................................................................................... 25
   2.3.4 In planta visualization of the endomembrane system and Cpr1 expression .......................................................... 25
   2.3.5 DNA Extraction .................................................................................................. 26
   2.3.6 Plant growth and inoculations ....................................................................... 27
   2.3.7 RNA extraction .................................................................................................. 28
   2.3.8 Transcriptome: Illumina RNA sequencing and data analysis ................... 29
   2.3.9 Heatmaps .......................................................................................................... 31
   2.3.10 Characterization of the MT Cpr1 allele ......................................................... 31
   2.3.11 Southern Blots ................................................................................................. 32
2.3.12 Transcriptome analysis to characterize transcripts of Cpr1 and other SPC genes \textit{in planta} ........................................................................................................... 32
2.3.13 Reverse transcription (RT)-PCR ................................................................. 33
2.4 Results .................................................................................................................... 35
  2.4.1 Genomics and comparative transcriptomics of the proposed secretory pathway in \textit{C. graminicola} ................................................................................. 35
  2.4.2 The SPC proteins in \textit{C. graminicola} ......................................................... 37
  2.4.3 Structure of the mutant and wild type Cpr1 alleles ................................. 38
  2.4.4 MT and WT Cpr1 transcripts \textit{in planta}. ................................................ 40
  2.4.6 RT-PCR amplification and sequencing of alternative transcripts from \textit{C. graminicola} mutant strain ................................................ 41
  2.4.7 Analysis of 3'UTR sequences of \textit{C. graminicola} MT and WT ......... 41
  2.4.8 \textit{In planta} analysis of the endomembrane system and expression and localization of CPR1 ................................................................. 42
2.5 Discussion ............................................................................................................. 44

Chapter 3 – Comparative genomics of the \textit{Colletotrichum graminicola} secretome and effectorome ................................................................................................................. 66
3.1 Overview ................................................................................................................ 66
3.2 Introduction .......................................................................................................... 67
  3.2.1 Effectors and fungal pathogenicity to plants ........................................ 67
  3.2.2 The role of effectors in \textit{Colletotrichum} .................................................. 68
  3.2.3 Identification of putative effectors ......................................................... 69
  3.2.4 Use of transcriptome data to identify candidate effectors ............... 70
  3.2.5 The role of effectors in non-host resistance ..................................... 71
  3.2.6 Rationale for identification of \textit{C. graminicola} candidate effectors ... 73
3.3 Material and Methods .......................................................................................... 75
  3.3.1 Fungal strains and genome data used in this study ......................... 75
  3.3.2 DNA extraction protocol ................................................................. 75
  3.3.3 Sequencing and assembly ................................................................. 76
  3.3.4 Gene annotation .................................................................................. 77
  3.3.5 Transcriptome data .............................................................................. 77
  3.3.6 Genome synteny .................................................................................. 77
  3.3.7 Identification of putative orthologous genes ........................................ 78
  3.3.8 Gene categorization ............................................................................ 78
3.3.9 Identification of candidate effector proteins........................................79
3.3.10 Identification of non-annotated putative effector proteins in C. graminicola.................................................................81
3.3.11 Expression analysis of the BAS3 homolog in the WT in vitro and in planta.................................................................82
3.4 Results .........................................................................................84
  3.4.1 Comparison of Colletotrichum genomes ....................................84
  3.4.2 Comparative analysis of Colletotrichum proteins .......................85
  3.4.3 Comparative genomics of Colletotrichum secretomes and effectoromes ........................................................................88
  3.4.4 Effector diversity among isolates ..............................................94
  3.4.5 Transcriptome analysis: expression of putative effector genes in MT vs WT C. graminicola in planta ..................................94
  3.4.6 Identification of non-annotated effectors ..................................99
  3.4.7 Expression of the C. graminicola BAS3 homolog in vitro and in planta. .................................................................100
  3.4.8 Identification of candidate effectors for future studies ............101
3.5 Discussion ..................................................................................102

Chapter 4 – Genomic and functional analysis of stress response in wild type and mutant strains of Colletotrichum graminicola in vitro and in planta ...... 124
4.1 Overview .....................................................................................124
4.2 Introduction ................................................................................124
  4.2.1. Fungal Response to Stress ....................................................124
  4.2.2 The Role of Stress Response in Pathogenicity .......................126
  4.2.3 Evidence that Cpr1 and the signal peptidase may play a role in stress response ..........................................................132
4.3 Material and Methods .................................................................133
  4.3.1 Strains and culture conditions .................................................133
  4.3.2. Pharmacological analysis of WT and MT stress responses in vitro .................................................................133
  4.3.3 In silico identification of C. graminicola stress response genes and pathways ...............................................................134
  4.3.4 Transcriptome analysis ...........................................................135
  4.3.5 Heatmaps .............................................................................136
  4.3.6 Expression analysis of selected stress response genes in the WT under conditions of stress in vitro and in planta ..................136
4.4 Results ................................................................................................................................. 140

4.4.1 The MT was more sensitive than the WT or MT-C strains to most stress-inducing chemicals and treatments ................................................. 140

4.4.2 About one-fifth of the C. graminicola genome encodes genes that are predicted to be involved in stress response ......................................... 140

4.4.3 Colletotrichum spp. encode homologs for a majority of the genes in the yeast secretory, oxidative, osmotic, and cell wall integrity stress response pathways ................................................................. 140

4.4.4 Presence of UPRE in stress genes ................................................................. 142

4.4.5 Genes involved in response to stress, and particularly to secretion stress, are highly expressed in planta ................................................................. 143

4.4.6 Patterns of differentially expressed genes in planta suggest that stress responses are most active early during infection by both the WT and the MT ..................................................................................................................... 144

4.4.7 Relatively few stress responsive genes are differentially expressed in planta .............................................................................................................. 145

4.4.8 Expression of selected stress response genes in vitro in response to chemically induced stress ................................................................. 148

4.4.9 Stress altered the structure of the MT Cpr1 transcripts in vitro compared with in planta ......................................................................................... 149

4.4.10 In planta visualization of the BiP stress response protein ............... 149

4.4.11 In vitro response of endomembrane system to stress chemicals 149

4.5 Discussion .......................................................................................................................... 151

Chapter 5 – Concluding remarks ......................................................................................... 177

Appendix I – Verification of transcriptome sequencing results .............................................. 181

AI.1 Background .................................................................................................................. 181

AI.2 Material and Methods ............................................................................................... 182

AI.3 Results .......................................................................................................................... 183

Appendix III – Optimization of the protocol for race tube growth assays ................. 192

AIII.1 Background ........................................................................................................ 192

AIII.2 Material and Methods .......................................................................................... 193

AIII.3 Results ...................................................................................................................... 194

References ............................................................................................................................ 200

Vita ........................................................................................................................................ 239
List of Figures

Figure 1.1 Disease cycle of *C. graminicola* ....................................................... 15
Figure 1.2 Sequential biotrophy of *C. graminicola* .............................................. 16
Figure 1.3 Growth in media (A) and *in planta* (B) of the three strains used in the experiments. ........................................................................................................ 17
Figure 2.1 Illustration of Cpr1 and the signal peptidase complex ......................... 50
Figure 2.2 Illustration of the pFPL Gateway vector construct ............................... 51
Figure 2.3 Degree of conservation of *Colletotrichum* secretory pathway based on yeast proteins ................................................................................................. 52
Figure 2.4 Model of the *Colletotrichum graminicola* secretory pathway together with normalized reads counts for each gene across growth stages . 53
Figure 2.5 Alignment of amino acid sequences of the four signal peptidase subunits...................................................................................................................... 54
Figure 2.6 Map of the MT 3’downstream region .................................................... 55
Figure 2.7 Southern Blots performed using MT DNA .............................................. 56
Figure 2.8 Transcript prediction for WT and MT Cpr1 gene using FGENESH57
Figure 2.9 Illustration of Cpr1 reads for WT and MT strains ............................. 58
Figure 2.10 Intron splicing in the four signal peptidase subunit transcripts ..... 59
Figure 2.11 Cpr1 intron pattern of WT and MT strains .................................. 60
Figure 2.12 Cpr1 transcript intron pattern of WT, MT and MT-C mycelia grown in Fries medium ..................................................................................................... 61
Figure 2.13 Prediction of the MT cpr1 transcript with the intron retained done by Geneious ........................................................................................................ 62
Figure 2.14 Sequencing of the 3’UTR of the WT and MT strains ....................... 63
Figure 2.15 WT and MT-HDEL tagged isolates .................................................... 64
Figure 2.16 Expression of CPR1-RFP *in planta* ............................................... 65
Figure 3.1 Bioinformatic pipeline for prediction of the *Colletotrichum* effectorome .................................................................................................................. 107
Figure 3.2 Synteny between *C. graminicola* and *C. sublineola* .................... 108
Figure 3.3 Amino acid similarity of orthologous proteins between *C. graminicola* and other sequenced *Colletotrichum* species .......................................... 109
Figure 3.4 Orthology by OrthoMCL ..................................................................... 110
Figure 3.5 Orthology by Reciprocal Best Hit ....................................................... 111
Figure 3.6 Carbohydrate-degrading enzymes (CAZymes) in *C. graminicola* (Cgram), *C. sublineola* (Csub) and *C. higginsianum* (Chig), separated by different classes. ...................................................................................................... 112
Figure 3.7 Blast results of proteins shared between *C. graminicola* and the other two species using NCBI database ......................................................... 113
Figure 3.8 Orthology among the small secreted proteins between all 3 *Colletotrichum* species ........................................................................................................ 114
Figure 3.9 Amino acid similarity of orthologous proteins between C. graminicola and the other two species, C. sublineola (A) and C. higginsianum (B) ................................................................. 115
Figure 3.10 Amino acid similarity of orthologous proteins between C. graminicola and the other two species, C. sublineola (A) and C. higginsianum (B) ...................................................... 116
Figure 3.11 Alignment of effector protein families. ................................................. 117
Figure 3.12 Sequence of motifs discovered on SSPs of C. graminicola using MEME. .................................................................................................................. 118
Figure 3.13 Comparison between secreted proteins present in the genome and secreted proteins present in differentially expressed RNAseq transcripts, showing the over-representation of secreted proteins in planta ...................................................... 119
Figure 3.14 Timing of the differentially expressed SSPs based on their homology to other Colletotrichum species ............................................................................. 120
Figure 3.15 Identification of non-annotated effectors in C. graminicola using FGENESH prediction programs .......................................................... 121
Figure 3.16 Expression of C. graminicola BAS3 on in planta and in vitro conditions ................................................................................................................... 122
Figure 3.17 Orphan candidate effectors highly expressed in the early stages of infection ........................................................................................................... 123
Figure 4.1 Illustration of the pFPL Gateway vector construct used to create protein fusions between C. graminicola BiP promoter region and the red fluorescent protein reporter gene. .......................................................... 157
Figure 4.2 Growth of C. graminicola strains during chemical stress .......... 158
Figure 4.3 Degree of conservation of Colletotrichum stress proteins versus their putative yeast homologs ........................................................................... 163
Figure 4.4 Stress pathway models ........................................................................ 164
Figure 4.5 Placement of response elements in the 5’ upstream region of genes involved in stress .................................................................................. 168
Figure 4.6 Heat maps .............................................................................................. 169
Figure 4.7 Analysis of stress-related genes present in the laser capture data with a logFC higher than 1 ..................................................................................... 171
Figure 4.8 RT-PCR with BiP, Hac and PDI during in vitro and in planta conditions ............................................................................................................... 172
Figure 4.9 Effect of stress on Cpr1 transcripts in vitro ....................................... 173
Figure 4.10 Transcript sequencing results of Cpr1 gene in the WT and MT strains ........................................................................................................... 174
Figure 4.11 Expression of BiP (GLRG_10629) chaperone reporter in vitro and in planta .............................................................. 175
Figure 4.12 WT-HDEL tagged isolate visualized in the confocal microscope ................................................................................................. 176
Figure AI. 1 Wild-type (WT) Cpr1 gene (green) is represented with flanking genes (pink).

Figure AI. 2 Technical replicates (Lanes), treatments and biological replicates (Rep) from the transcriptome sequencing project.

Figure AI. 3 Transcriptome reads showing matches to the Cpr1 gene and the two flanking genes in the WT (blue) and the MT (red) strains.

Figure AI. 4 Hypothesized mutant sequence with reads matching the plasmid regions.

Figure AI. 5 Map of pAN56-1-sGFP-HDEL plasmid.

Figure AI. 6 Correlation analysis of fold changes by RNA sequencing and RT-PCR.

Figure AI. 7 Corrected map of pCB1636.

Figure AIII.1 Variability in radial growth during chemical stress.

Figure AIII.2 Race tubes growth.

Figure AIII. 3 Alignment between S. cerevisiae Kar2/BiP (YJL034W) and C. graminicola homolog GLRG_10629.

Figure AIII. 4 Primers and sequencing results of the promoter region used to produce a reporter construct for BiP using Gateway.
List of Files

EsterBuiate_Dissertation_Files.xlsx File size 3.89 MB
Chapter 1

*Colletotrichum graminicola*, the causal agent of maize anthracnose

1.1 Overview

Anthracnose stalk rot, caused by the fungus *Colletotrichum graminicola* (Ces.) Wilson, is one of the most important diseases of maize and causes significant losses worldwide. The most effective control for the disease is the use of resistant cultivars, but the success of this approach depends not only on the host genetics, but also on the plant physiological state and environmental conditions that can change from year to year.

*C. graminicola* is a hemibiotroph which initially establishes an asymptomatic biotrophic infection that is followed later by a switch to necrotrophy when symptoms are produced. Our laboratory produced a mutant strain of *C. graminicola* several years ago that is completely nonpathogenic. The mutation is in a gene encoding one of the components of the signal peptidase complex, known to be important for protein processing and secretion. This mutant fails to establish biotrophy. Our long-term goal is to understand the function of the mutated gene and its relation to disease development. The goal of my dissertation research was to use a genomic and bioinformatics approach to address several hypotheses related to understanding the difference in the ability of the wild type (WT) and the mutant (MT) to establish biotrophy.

1.2 The economic impact of maize anthracnose

Maize (*Zea mays* L.) is one of the most important cereal crops in the world, and is the most valuable crop cultivated in the United States (USA), which is the biggest producer. In 2012, worldwide production of maize was 872 million tons, more than any other cereal. The USA alone produced around 31.4% of the total (FAO, 2014). Besides being a staple food in some regions of the world, maize can be used for the production of oil, syrup, alcohol, livestock feed, and more recently for biofuel. Maize stalks can also be a source for cellulosic ethanol production. The increase in production of ethanol biofuel from maize in the USA
has resulted in corresponding increases in the value of the crop and in the acreage grown. Continuous cropping has become more common. One of the major factors limiting maize production is disease. It is estimated that diseases cause losses from 2 to 15% annually, and anthracnose in particular is a disease of worldwide importance on maize (Balint-Kurti and Johal, 2009; White, 1999).

Anthracnose can affect all parts of the plant, but it is most commonly associated with leaf blight (ALB) and stalk rot (ASR). ASR is by far the more important of the two. Maize stalk rots can be caused by several fungi other than *C. graminicola*, including *Fusarium graminearum* and *Stenocarpella maydis* (Denti and Reis, 2003; Ribeiro et al., 2005). In the USA, *C. graminicola* is considered to be the most important stalk rot pathogen although it was considered a minor problem prior to 1970. In the early 1970s several severe ASR epidemics occurred that caused lodging and in some cases complete crop loss (Bergstrom and Nicholson, 1999) and now ASR is the most common and damaging of the stalk rots (Denti and Reis, 2003; Ribeiro et al., 2005). Although there is relatively little information available on anthracnose disease severity, Jirak-Peterson and Esker (2011) reported disease levels in Wisconsin of around 5% during a two year trial, while work done by Costa et al. (2010) recorded disease severities ranging from 30 to 60% in Brazil. Both studies suggested that the hybrid chosen was an important factor for the occurrence of anthracnose stalk rot.

ASR disease management involves deployment of resistant cultivars, control of the European corn borer and corn rootworms, and cultural practices such as balanced soil fertility, stress reduction and proper planting rates (Bergstrom and Nicholson, 1999). Occasionally, fungicides are used, especially in maize seed production fields in Brazil. Increases in disease pressure and in grain prices have led to more fungicide use for maize production in the last decade (Wise and Mueller, 2011). Although the levels of control achieved are not large, the small increase in production caused by strobilurin fungicides (due to delay of plant senescence) is an incentive for farmers to apply the chemical (Byamukama et al. 2013; Vincelli et al. 2013). Anthracnose disease levels depend largely on plant maturity and environmental conditions, which can vary from year to year, making it hard to develop a management strategy based
solely on genetic resistance. Thus, chronic losses of between 5-10% are believed to occur annually, and epidemics resulting in losses of up to 100% occur sporadically and without warning (Frey et al. 2011; Bergstrom and Nicholson 1999). A better understanding of how the pathogen infects and colonizes maize is essential for more consistent and effective control of the disease (Ceasar and Ignacimuthu, 2012; Tripathi et al. 2014).

1.3 Disease cycle of anthracnose

The causal agent of maize anthracnose is the hemibiotrophic fungus *Colletotrichum graminicola* (Ces.) G.W. Wilson (Sutton, 1980). Until recently, *C. graminicola* was thought to infect sorghum and other grasses in addition to maize, but isolates from these other grass species are now known to belong to different, related species (Vaillancourt and Hanau, 1992; Crouch et al. 2009). *C. graminicola* differs in morphology from the closely related species that infects sorghum (*C. sublineola*): attempts to cross them did not produce progeny, and results of DNA analyses demonstrated they are reproductively isolated sibling species (Crouch et al., 2009; Vaillancourt and Hanau, 1992). Wheeler et al. (1974) reported that maize isolates could infect sorghum in a growth chamber when the inoculated plants were incubated for 24 hours at 100% moisture. Venard and Vaillancourt (2007b) showed that *C. sublineola* could complete its life cycle in maize stalks, although colonization was confined to the epidermal cells. However, the majority of reports suggest that the maize and sorghum isolates of *Colletotrichum* are host-specific, and cross-inoculation has never been reported to occur in the field (Jamil and Nicholson, 1987; Snyder et al. 1991; Torres et al. 2013).

The anthracnose disease cycle (Figure 1.1) starts with acervuli produced on overwintering crop debris on the soil surface, from which conidia are disseminated by rain splash to nearby maize leaves or stalks. Production of secondary inoculum happens in the contaminated tissues, usually on the lower leaves. Many secondary infection cycles can happen in the same season (Bergstrom and Nicholson, 1999; Crouch et al., 2014). The pathogen can overwinter in the residues as a saprophyte and starts sporulating in the spring, where it serves as a source of primary infection during the next season (Naylor
and Leonard, 1977). Reports showed that burying plant residue from the previous year may result in more stalk rot (Jirak and Esker, 2009) and that the incidence of disease was 78% higher when corn was planted continuously than when rotated with soybeans (Jirak-Peterson and Esker, 2011). If infested cornstalks were buried at least 10-14 cm deep for eight months, only 2% of them produced sporulating acervuli, whereas sporulation occurred on all of the stalks left on the soil surface (Lipps 1983).

In highly susceptible maize cultivars, it was observed that planting infected kernels resulted in root decay (Warren and Nicholson, 1975). Bergstrom and Nicholson (1999) suggested that the pathogen could spread to stalk tissues from root infections via the vascular tissue. The pathogen can be readily recovered from isolated vascular bundles (Bergstrom and Bergstrom, 1987). Sukno et al. (2008) reported the systemic spread of the pathogen from infected roots through the xylem, but Venard and Vaillancourt (2007b) did not find evidence for systemic movement of the pathogen in vascular bundles, although they did find that the hyphae could colonize the associated nonliving fiber cells. The fungus was able to move quickly along the vascular bundles via these fiber cells, and emerge to attack cells far from the infection point (Venard and Vaillancourt, 2007a). The pathogen grows through the bundle sheath and fiber cells surrounding vascular bundles asymptotically (Mims and Vaillancourt, 2002; Venard and Vaillancourt, 2007a, 2007b).

1.4 Hemibiotrophy and its role in the disease cycle in *Colletotrichum*

*C. graminicola*, like other *Colletotrichum* species, is a hemibiotroph. Conidia adhere to the plant surface, germinate and produce a melanized appressorium that facilitates penetration of the plant cell wall, a process helped by the production of cell wall degrading enzymes (Bergstrom and Nicholson, 1999; Nicholson et al.1976). Penetration occurs via the penetration pore, a circular zone at the base of the appressorium that is not melanized from which the penetration peg emerges. The epidermal cells are invaded by primary hyphae that grow to colonize neighboring cells. The primary infection hyphae are multinucleate, swollen, and irregularly shaped and they are surrounded by a membrane that separates them from the living host cells (Bergstrom and
Nicholson, 1999; Mims and Vaillancourt, 2002; Venard and Vaillancourt 2007a,b). Growth of the primary hyphae from cell to cell is via narrow cell connections through extremely thin hyphal connections (Venard and Vaillancourt, 2007a,b). Newly colonized cells are still alive while the cells behind the infection front quickly become granulated and die, a phenomenon that we have called sequential biotrophy (Figure 1.2). Secondary hyphae, which are narrower, uninucleate, and not surrounded by a membrane, are produced as branches from the primary hyphae in the invaded cells after the cells die and the pathogen enters its necrotrophic phase of growth (Venard and Vaillancourt, 2007a,b, Torres et al., 2013). At some point after the pathogen switches to necrotrophy, host cell walls become degraded and symptoms of anthracnose become evident. Similar processes occur both in leaves and in stalks (Cadena-Gomez and Nicholson, 1987; Mims and Vaillancourt, 2002; Tang et al. 2006; Venard and Vaillancourt, 2007a, 2007b).

The issue of whether the C. graminicola had a biotrophic phase or not was raised by O’Connell et al. (2000). Politis and Wheeler (1973) reported that the plasma membrane remained intact and surrounded the base of the appressorium, suggesting a biotrophic stage, but they also saw disrupted plasma membrane in some cells, which they attributed to the tissue preparation for microscopy. The issue was settled by Mims and Vaillancourt (2002) when they observed the presence of a membrane surrounding biotrophic hyphae inside living host cells. This has since been further supported by work done by two independent research groups that confirmed plasmolysis in maize cells containing C. graminicola biotrophic hyphae (Tang et al. 2006; Torres et al., 2013).

Besides sequential hemibiotrophy, which is typical of C. graminicola and C. sublineola, there are two other types of Colletotrichum hemibiotrophy (Crouch et al. 2014). In C. higginsianum, primary hyphae occur only in the epidermal cells, and then these produce secondary hyphae that kill the neighboring cells in advance, prior to necrotrophic invasion. In C. orbiculare, the primary hyphae colonize several layers of neighboring cells biotrophically, but at some point, like C. higginsianum, this fungus also makes a complete switch to necrotrophic hyphae that begin to kill cells in advance of colonization. Thus, the main feature
that differentiates hemibiotrophy in \textit{C. graminicola} and \textit{C. sublimeola} from hemibiotrophy in other \textit{Colletotrichum} species is the persistence of the biotrophic phase at the colony edges, and the resulting co-existence of necrotrophy and biotrophy in expanding lesions (Torres et al., 2013). In all three types of hemibiotrophy, the biotrophic step appears to be critical for the establishment of a successful infection in the living host (Crouch et al., 2014).

To successfully infect a host, hemibiotrophic and biotrophic fungi deploy secreted effector proteins that promote virulence, and allow the establishment of biotrophy (Dou and Zhou, 2012; Kleemann et al., 2012; Wit et al. 2009). Necrotrophic fungi often produce secondary metabolites that act as phytotoxins, leading to death of the host cells and allowing them to colonize the dead cells (Vleeshouwers and Oliver, 2014). The genomes of \textit{Colletotrichum} species encode expanded repertoires of secreted proteases, CAZymes, secondary metabolites, and secreted effector proteins, and the primary hyphae seem to function mainly as secretory cells for these products (O'Connell et al., 2012; reviewed by Crouch et al. 2014). The effectors and secondary metabolites produced during biotrophy are assumed to suppress plant defenses and cell death, at least temporarily, and help to establish compatibility (Kleemann et al., 2012; O'Connell et al., 2012; Rafiqi et al. 2012).

During the interaction with the plant it is thought that the fungus encounters various stresses, including oxidative, nutritional and secretion stress, to which it must adapt (Torres, 2010). Maize plants react to fungal invasion by the production of various defensive compounds and by the construction of lignified papillae in the outer walls of epidermal cells (Bergstrom and Nicholson, 1999; Mims and Vaillancourt, 2002; Politis and Wheeler, 1973). Cells surrounding infected cells within lesions are also fortified by lignification, as well as by the deposition of polymers that make cell wall more resistant to CWDE and induction of chitinases that attack the fungal cell wall (Cadena-Gomez and Nicholson, 1987). None of these measures seems to prevent the establishment of \textit{C. graminicola} in a susceptible host (Mims and Vaillancourt, 2002; Politis and Wheeler, 1973).
1.5 Resistance to maize anthracnose

Maize anthracnose is managed primarily by planting resistant hybrids. The resistance to *C. graminicola* currently utilized in commercial hybrids is primarily quantitative, aka. partial, resistance. This type of resistance has been quite effective: thus, current hybrids are much more resistant than those that were commonly planted in the early 1970s, when several major epidemics of anthracnose occurred (Bergstrom and Nicholson, 1999; Leonard and Thompson, 1976). However, these quantitative sources can fail if populations of the pathogen are high (as occurs in continuous maize cropping, or with no-till or reduced tillage management systems), and/or if the plants are stressed (as occurs during drought, significant insect or pathogen damage, or high plant populations that result in nutritional or light stress). These quantitative resistance sources also become less effective during flowering and grain fill, which can lead to late-season stalk damage and lodging (Dodds and Schwechheimer, 2002). It is complicated and time-consuming for breeders to introduce new quantitative resistance sources by traditional techniques; thus, we are unable to react quickly to the shifts that occur in the pathogen population in response to currently deployed sources of resistance.

A major-gene source of resistance for anthracnose stalk rot, *Rcg1* (Resistance to *Colletotrichum graminicola*), was described by Frey et al. (2011). *Rcg1* is a complex locus that contains two LRR-type R genes, both of which are required for the expression of resistance (Frey et al. 2011). Only about 6% of the maize lines that are used for breeding in North America contain *Rcg1* (Broglie et al. 2011). Inbred lines containing the *Rcg1* locus were highly resistant to stalk rot disease and delivered a higher yield when compared with near isogenic lines that did not contain the resistance genes that were exposed to the same amount of disease pressure (Frey et al. 2011). Pioneer is currently developing commercial hybrids containing *Rcg1*. Other major-gene sources of resistance for ASR and ALB have been described, but none of these are currently in commercial development, to my knowledge (Matiello et al. 2012; Badu-Apraku et al. 1987; Badu-Apraku personal communication). Resistance to ALB is not always correlated with ASR resistance, and resistance to anthracnose usually does not confer resistance to other stalk rot pathogens, e.g. *Stenocarpella*
maydis and Gibberella zeae (Nyhus et al. 1989; Sweets and Wright, 2008; Zuber et al. 1981). Additionally, there is concern that wide deployment of major gene sources of resistance could lead to selection of pathogenic races and “boom-bust” epidemics, similar to those that occur in sorghum infected by the closely related pathogen C. sublineola. Unlike maize anthracnose sorghum anthracnose is managed mainly by the use of vertical (aka. major gene, or qualitative) resistance, and the C. sublineola population contains a very large number of different races (Casela et al. 2004; Prom et al. 2012; Valério et al. 2005). Until recently, races were not believed to exist in the population of C. graminicola (Forgey et al. 1978; Nicholson and Warren, 1981), but a study last year confirmed the existence of races for the first time, occurring rarely in the population of C. graminicola in Brazil (Costa et al., 2014).

Long-term management of anthracnose by using either quantitative or qualitative sources of resistance developed by traditional breeding will remain challenging. A better understanding of the mechanisms that are critical for infection by C. graminicola might suggest novel targets for chemical or biotechnological therapies that could provide a more durable and effective solution.

1.6 The Cpr1 mutant of C. graminicola

A non-pathogenic C. graminicola MT was produced several years ago in our laboratory by restriction enzyme-mediated insertional (REMI) mutagenesis (Thon et al. 2002; Thon et al. 2000). This MT is normal in culture, except for a slightly reduced growth rate, but it is unable to cause any symptoms in either maize leaves or stalks (Figure 1.3, Thon et al., 2002; Mims and Vaillancourt, 2002; Venard and Vaillancourt, 2007a,b). The insertional mutation occurred in the 3’UTR region of a homolog of the yeast Spc3 gene, 19 bp downstream from the stop codon. The gene was named Cpr1, for Colletotrichum pathogenicity related gene 1. Transformation of the MT with a sub-clone containing the Cpr1 gene resulted in complementation, demonstrating that the mutation in Cpr1 was responsible for the nonpathogenic phenotype (Thon et al., 2002). The MT is leaky, apparently producing a small quantity of normal transcript, which seems
to be enough for near-normal in vitro growth but not for a successful plant infection (Thon et al., 2002).

The yeast Spc3 gene encodes a non-catalytic component of the endoplasmic reticulum (ER)-localized signal peptidase complex (SPC), which is essential for protein processing and secretion (Fang et al. 1997). In yeast the SPC consists of four subunits. Sec11p and Spc3p are essential for signal peptidase activity in yeast, and the deletion of either of the genes encoding these proteins is lethal (Fang et al., 1997; Paetzel et al. 2002). Sec11p is required for signal peptide cleavage and signal peptidase-dependent protein degradation. The function of Spc3 is currently unknown, but in yeast it interacts with Sec11p, as demonstrated by co-immunoprecipitation experiments (Fang et al., 1997; VanValkenburgh et al. 1999). The other subunits of the yeast SPC are Spc1p and Spc2p. These perform auxiliary and non-redundant roles, and they are both non-essential for cell growth and enzyme activity (Fang et al., 1997). Spc2 was found to be necessary for growth in high temperatures in yeast (Mullins et al. 1996).

Cytological and ultrastructural observations of infection in leaves of susceptible maize plants demonstrated that the MT produces normal appressoria and penetrates epidermal cells, and also mesophyll and bundle sheath cells to a very limited degree (Thon et al., 2000; Mims and Vaillancourt, 2002). Both the MT and the WT caused cell death, but the MT produced very few dead cells, and failed to switch to necrotrophic growth in mesophyll cells (Thon et al., 2000; Mims and Vaillancourt, 2002).

A subsequent, more detailed cytological study in living maize leaf sheaths confirmed normal production of appressoria by the MT, but revealed that the production of primary invasive hyphae is delayed by approximately 24 hours relative to the WT strain. The MT was mostly unable to progress from the first invaded cell to establish the normal biotrophic phase of development (Torres et al., 2013). The authors reported that when the host cells were killed, the MT colonized the tissues at the same rate as the WT and progressed to sporulation, showing it has the ability to grow saprophytically in the maize tissues. Even more interesting, these authors showed that when the MT and WT strains are inoculated close together, the MT is able to grow normally. This suggested the
hypothesis that the WT is able to produce some kind of a diffusible substance or signal that renders nearby host cells receptive to fungal colonization, and that the MT is unable to produce these substances or signals.

1.7 What is the role of Cpr1 in C. graminicola pathogenicity?

To successfully infect a plant, fungal pathogens must process and secrete many proteins that are necessary for inducing susceptibility, adapting to the stressful plant environment, and utilizing plant tissues for nutrients (Coaker, 2014; de Jonge et al. 2011; Dickman and de Figueiredo, 2013; Gan et al., 2012; Kamoun, 2007; Kombrink and Thomma, 2013; Tang et al., 2006; Torres et al., 2013; Valent and Khang, 2010). The SPC is critical for processing and secretion of proteins. The fact that the MT strain can grow when the plant tissue is dead, or when it is grown in close proximity to the WT strain in living sheath tissues, (Torres et al., 2013) suggests that it is deficient specifically in its ability to induce susceptibility and/or adapt to conditions in a living host that is actively defending itself.

Bacterial Type 1 signal peptidases have been directly implicated in pathogenicity. Signal peptidases of Listeria monocytogenes, Escherichia coli and Staphylococcus aureus occur as a series of paralogs, some of which are specifically involved in the secretion of virulence factors (Choo et al. 2008; Kavanaugh et al. 2007; Raynaud and Charbit, 2005). In Streptococcus pneumoniae, mutation of the signal peptidase reduced pathogenicity in a mammalian host (Khandavilli et al. 2008).

I propose two hypotheses: a) the MT can’t adapt to the stressful environment resulting from active plant defenses, and/or b) it can’t produce or secrete proteins necessary for induction of susceptibility. Unfortunately, even though it appears to be universally conserved in eukaryotic organisms, very little is known about the precise role of the CPR1 protein and its homologs in SPC function. We know that it is an essential protein, because knockouts are lethal in yeast and Candida albicans, and presumably also in C. graminicola where attempts to obtain a viable knockout strain failed (De la Rosa et al. 2004; Fang et al., 1997; Thon et al., 2002). The C. graminicola MT is unique because the leaky insertional mutation results in a conditional effect during in vitro versus in
plantation growth. Understanding the specific nature of this mutation and its role in establishment of biotrophy could lead to a novel means to disrupt and prevent biotrophic establishment, and thus potentially to new and highly effective disease management tools.

1.8 Hypotheses and objectives of this dissertation

My dissertation is divided into three chapters, each focused on a different hypothesis related to the putative role of Cpr1 in pathogenicity. I have utilized genomics and bioinformatics approaches for my dissertation research. Our laboratory, with our collaborators, sequenced the C. graminicola strain M1.001 (O’Connell et al. 2012), a second C. graminicola strain (M5.001), and a C. sublineola strain (CgSl1). We also sequenced the transcriptomes of the M1.001 WT strain during different stages of growth in planta (O’Connell et al., 2012) and of the cpr1 MT strain at comparable stages.

Genomics and transcriptomics provide essential bases for future work. These are fast, high-throughput, and relatively inexpensive techniques. Analysis of the genomes and transcriptomes helps to identify gene candidates for future functional analyses of their potential role in pathogenesis. The transcriptome provides data that can be used to surmise regulatory interactions, including coordination of signaling pathways and gene clusters. Comparative genomics can reveal differences between closely related pathogen strains that vary in virulence, and identify candidate genes that are potentially important for host or cultivar specificity (Bhadauria et al. 2007; Nemri et al., 2014).

I am aware that genomic data also have limitations. Although some studies show a good correlation between transcriptomes and proteomes (Barker et al. 2012), others suggest they are only weakly correlated (Haider and Pal, 2013; Washburn et al., 2003). Levels of proteins involved in translation or stress response, in particular, were poorly correlated with transcript levels in pathogenic Staphylococcus (Carvalhais et al. 2015). Transcript structure, e.g. 3’UTR sequence, can have a significant impact on translation efficiency and transcript stability, and thus on protein levels (Horgan and Kenny, 2011). Many transcripts occur in very low abundance, or have short half-lives, and thus are under-represented in a transcriptome study. Because the plant to fungal tissue
ratio is high, fungal transcripts are much less abundant than plant transcripts in mixed extracts from infected plant tissues, thus many fungal transcripts may be overlooked in these combined transcriptome studies. In spite of these limitations, genomics and transcriptomics represent an essential starting point for developing new hypotheses about *C. graminicola* pathogenicity that can be tested in the future.

1.8.1 Hypothesis 1: the mutation in the *Cpr1* gene results in abnormal splicing of the transcript sequence.

Northern blots suggested that the MT strain made very little of the normal transcript, and instead produced a variety of aberrant transcripts that were both longer and shorter than normal *in vitro* (Thon et al., 2002). These data suggested the hypothesis that the *Cpr1* transcript sequence is altered by differential splicing during some phases of development in the WT and/or the MT, and that this leads to changes in the quality or quantity of the CPR1 protein. The 3'UTR sequence, where the *Cpr1* mutation occurred, has been implicated in regulating transcript splicing and polyadenylation, nuclear export, transcript stability, translation efficiency, and mRNA targeting (reviewed by Grzybowska et al. 2001). In chapter 2, I addressed this hypothesis by investigating whether transcript sequences differed in the MT and WT strains *in vitro* and during development *in planta*.

1.8.2 Hypothesis 2: The MT is nonpathogenic because it fails to secrete effectors that are necessary for infection and establishment of biotrophy in maize.

The literature suggests that secreted proteins, and particularly a class of highly divergent, small secreted proteins known as effectors, are very important in the establishment of infection by plant pathogenic fungi. A study with *Medicago truncatula* demonstrated that a plant orthologue of *Cpr1* was essential for establishment of nodule formation (Wang 2010), apparently because it was specifically required for secretion of small protein effectors necessary for conversion of the bacteria to bacteroids (Van den Velde et al., 2010). Because the MT is affected in a putative component of the secretory pathway, and because cytological evidence suggests that the MT may fail to produce
diffusible factors necessary for establishment of biotrophy in the living host, it is logical to investigate the nature of the *C. graminicola* secretome and its expression *in planta*. This is a first step toward testing the hypothesis that an inability to produce important secreted proteins is responsible for the MT phenotype, by identifying the most likely candidates for those critical proteins. Chapter 3 contains the results of my comparative bioinformatics analysis of the genomes and secretomes of *C. graminicola* and of the very closely related fungus *C. sublineola*, which is not a pathogen of maize. My rationale for this study was that idea that secreted proteins that are directly relevant to the establishment of biotrophy in maize are likely to be under strong selection pressure, and thus comparison with *C. sublineola*, which is closely related but fails to establish a successful biotrophic interaction with maize, may identify the most likely candidates for effectors with this role. Additionally, I reasoned that transcriptome analysis of the WT vs MT would help to identify effectors that are expressed at the right time (that is, early) to be involved in biotrophic establishment. My goal for this work was to identify a list of the most promising such candidate effectors that can be tested by future researchers in functional studies, and to test my prediction that effectors that are most divergent between *C. graminicola* and *C. sublineola* would also be those that were expressed early, during the establishment of biotrophy.

1.8.3 Hypothesis 3: the MT is nonpathogenic because it cannot adapt to secretion stress and/or other stresses that are encountered *in planta*.

There is evidence in the literature that plant tissues that are actively defending themselves produce a stressful environment for the pathogen (Dou and Zhou, 2012; Lowe and Howlett, 2012; O’Connell and Panstruga, 2006; Vleeshouwers and Oliver, 2014). The MT is able to grow normally in maize tissues that are not alive and actively defensive (Torres et al., 2013). The question arises, when *C. graminicola* establishes a biotrophic infection, does it induce plant defenses and activate its stress response pathways? And might the MT be deficient in these activities? Some evidence to suggest a role for Cpr1 in stress response came from a study on responses to secretion stress by *Aspergillus niger*. Guillemette et al. (2007) reported that the Cpr1 homolog in *A. niger* was transcriptionally up-regulated by 2-fold during chemically-induced secretion stress. Moreover, it
was post-transcriptionally up-regulated by 7-fold in response to the stress, more than any other gene in their study! The *Sec11* orthologue and the other components of the SPC were not similarly up-regulated in *A. niger*. This observation suggested the hypothesis that *Cpr1* could play an important and specific role in helping the pathogen deal with secretion stress or other stresses occurring during the establishment of infection. To address this hypothesis, in chapter 4 I tested whether the MT is deficient in the ability to adapt to stress *in vitro*, and I investigated expression of genes involved in stress response in the WT versus MT *in vitro* and *in planta*. I also expected to receive some insights from this work into the types of stress that are being experienced by the WT during various phases of development *in planta.*
Figure 1.1 Disease cycle of *C. graminicola*. Fungus overwinter in plant stalks (A), producing primary inoculum during spring (B and C). Spores are disseminated by rain (D) and infect corn plants (E). Pathogen produce secondary inoculum on infected stalks (F) and leaves (G). Figure B courtesy of Dr. Lisa Vaillancourt. Figure 4 from USDA Natural Resources Conservation Service (http://conservationdistrict.org/2015/the-power-of-a-raindrop.html).
Figure 1.2 Sequential biotrophy of *C. graminicola*. The pathogen penetrates the epidermal cell (1) via an appressorium (asterisk). The primary hyphae branches and sequentially infects other host cells (2, 3 and 4). Granulation of the plant cytoplasm is observed in the first invaded cell.
Figure 1.3 Growth in media (A) and *in planta* (B) of the three strains used in the experiments: WT (wild-type strain), MT (mutant strain) and MT-C (complemented strain). The whitish flecks on the maize leaves are caused by thrips damage, not disease.
Chapter 2

Transcriptional analysis of *Cpr1*, and other components of the signal peptidase complex and secretory pathway, in wild type and mutant strains of *Colletotrichum graminicola*

2.1 Overview

Our laboratory produced a MT strain of *C. graminicola* that is nonpathogenic to maize. The mutation is in a gene called *Cpr1*. The CPR1 protein is a homolog of one component of the microsomal signal peptidase complex, which comprises the first step in the canonical secretory pathway in eukaryotic organisms. The precise function of the conserved CPR1 subunit of the eukaryotic signal peptidase is unknown, and there is very little published research. However, it appears to be universally conserved in eukaryotes, and it is essential for viability in yeast. *C. graminicola* has only a single copy of this essential housekeeping gene. Lack of the CPR1 protein should be highly debilitating. How can we explain why the MT is apparently normal in culture, and during early stages of pathogenicity, but is deficient specifically in the ability to establish biotrophy? The conditional nature of the mutation could be related to qualitative or quantitative changes in the transcript and/or the protein. The MT has an insertion in the 3’UTR of the *Cpr1* gene, and the 3’UTR is known to regulate various post-transcriptional processes, including transcript stability, transcript splicing, and rates of translation. One possibility is that *C. graminicola* needs more of the CPR1 protein specifically during biotrophy, and that the MT is unable to produce adequate amounts to support this transition due to transcript instability, aberrant splicing, or reduced translation rates. Another possibility is that the WT produces different proteins via alternative splicing, one of which is specifically functional *in planta*, and that the MT is unable to undergo this alternative splicing. In this chapter I explore some of these possibilities by characterizing the *Cpr1* transcripts produced by the MT, WT, and complemented MT (MT-C) strains *in planta* and *in vitro*. To accomplish this, it was first necessary for me to characterize the insertion in the 3’ UTR of the MT strain in detail, something that had not been done previously. I also conducted
a broader bioinformatic and comparative transcriptomic analysis of other putative conserved components of the secretory pathway in the MT and WT in planta, in order to address some alternate hypotheses about the conditional nature of the Cpr1 mutation, and to determine whether there were differences in the expression of other secretory pathway genes between the MT and WT strains.

2.2 Introduction

In an effort to identify novel C. graminicola genes critical for pathogenicity, our laboratory produced a collection of random mutants by using the restriction enzyme-mediated insertional (REMI) mutagenesis technique, and screened them for loss of virulence on maize leaves and stalks (Thon et al., 2000; Thon et al., 2002). The REMI mutagenesis experiment was done using EcoRI. A plasmid called pCB1636 (Sweigard et al., 1997), that had also been linearized with EcoRI, was inserted randomly into the genome (Thon et al., 2000; Thon et al., 2002). One mutant identified from this study was completely non-pathogenic to living maize leaves and stalks, but capable of near-normal growth in culture (Thon et al., 2002; Mims and Vaillancourt 2002). This mutant is also able to complete its life cycle in killed maize tissues, and it germinates and penetrates living maize epidermal cells normally (Torres et al., 2013). However, it ultimately fails to establish a successful biotrophic infection once it enters the host cells (Torres et al., 2013). The mutation is in a gene we called Cpr1 (Colletotrichum pathogenicity related gene 1). Introduction of a subclone containing the wild-type Cpr1 gene at an ectopic site complemented the mutant and restored pathogenicity (Figure 1.3C; Thon et al., 2002).

The REMI plasmid sequence was inserted into an EcoRI site 19 base pairs downstream from the predicted stop codon of Cpr1, interrupting the 3’UTR (untranslated region) of the gene (Figure 2.1.A) (Thon et al., 2000; Thon et al., 2002). It was proposed that one complete and one incomplete copy of the REMI plasmid were integrated in tandem at this site. This was determined by a combination of Southern hybridization analysis and sequencing the upstream end and flank of the insertion, which were rescued in Escherichia coli.
The downstream end and flank of the insertion were not successfully rescued.

A postdoctoral researcher in our laboratory, Dr. Eunyoung Park, used 3’ and 5’ RACE to confirm the predicted sequence of the complete WT Cpr1 transcript \textit{in vitro} (Figure 2.1.A, E. Park unpublished data). However, Dr. Park was unsuccessful when she tried to use a similar approach to characterize the MT transcript. A graduate student in our laboratory, Dr. Maria Torres, used real-time quantitative reverse-transcription (RT)-PCR to demonstrate that the total amount of \textit{Cpr1} transcript produced by the MT was the same as the WT or MT-C strains \textit{in planta} (Torres, 2013). Northern blots published by Thon et al. (2002) indicated that the mutant had a severe reduction in the amount of normal \textit{Cpr1} transcript \textit{in vitro}, compared with the WT and MT-C strains, but that it also produced a variety of additional transcript species that were both larger and smaller than the normal size.

The CPR1 protein is homologous to Spc3p in \textit{Saccharomyces cerevisiae}. Spc3p is one subunit of the signal peptidase complex (SPC), an integral endoplasmic reticulum (ER) membrane protein complex that is important for cleaving signal peptides of proteins destined for the secretory pathway (Figure 2.1.B). The SPC in \textit{S. cerevisiae} is comprised of four subunits: Sec11p, Spc1p, Spc2p, and Spc3p. Sec11p and Spc3p appear to be universally conserved in all eukaryotic organisms (Paetzel et al. 2002; Antonin et al. 2000; Liang et al. 2003). Sec11p is the catalytic subunit, and it is essential for signal peptide cleavage and for viability (Böhni et al., 1988). The residues that are directly involved in the catalytic function in Sec11p appear to be conserved across a range of organisms studied (VanValkenburgh et al. 1999). Spc1p and Spc2p are not essential for cell growth in yeast, and they appear to have auxiliary roles. Spc2 increases the enzymatic activity of the complex, and is also thought to interact with proteins from the translocon pore and facilitate cleavage of the nascent polypeptide chain (Antonin et al. 2000). Spc2p has been shown to be important for optimal growth and function at high temperatures (Mullins et al. 1996). The Spc3p subunit, like Sec11p, is essential for viability and for activity of the complex, although it does not appear to have a direct catalytic function (Fang...
et al. 1997). Work with temperature sensitive Sec11 and Spc3 yeast mutants has shown that they accumulate misfolded proteins (Meyer and Hartmann 1997; Böhni et al. 1988). This results in activation of the conserved Unfolded Protein Response (UPR) (aka. secretion stress) pathway (Travers et al. 2000), which functions to maintain viability when protein transport is interrupted, and the ERAD (ER-associated degradation) pathway, that works to remove misfolded proteins from the cell. Work by Fang and collaborators (1997) shows that overexpression of Spc3p in yeast can suppress the Sec11 mutation, but that the opposite is not true. Although the precise function of Spc3p is unknown, it is proposed that it serves to stabilize the catalytic Sec11p subunit of the SPC (Meyer and Hartmann 1997).

Lack of the CPR1 protein in C. graminicola should be highly debilitating, yet the MT is nearly normal in culture and during the early stages of pathogenicity. How can we explain its deficiency specifically in the ability to establish biotrophy? In some organisms, there is more than one paralog of some subunits of the SPC. For example, in mammals there are two homologs of the catalytic Sec11 subunit, SPC18 and SPC21. Their sequences are extremely similar (Shelness and Blobel 1990) and although they have overlapping substrate specificity, they show different efficiencies in processing the same transcript (Liang et al. 2003).

In the model legume M. truncatula, there are two proteins, DNF1L and DNF1, both annotated as homologs of the Spc3p subunit. The two paralogs share 82% identity at the amino acid level. The first one is important for normal plant function, while the other is expressed only in nodule cells infected with Rhizobium, and is essential for nodulation (Wang et al. 2010). It appears that this second protein functions specifically in processing of plant secreted proteins that are needed to induce transformation of Rhizobium bacteria into bacteroids (Van de Velde et al. 2010). Although the paper by Thon et al. (2002) included Southern blot evidence for a potential paralog of Cpr1, subsequent analyses of the sequenced C. graminicola genome have failed to confirm this.

In the absence of multiple isoforms, post-transcriptional or post-translational regulation could explain the conditional behavior of the cpr1 mutant.

There is evidence in the literature for post-transcriptional regulation of genes encoding components of the SPC. Guillemette et al. (2007) reported that the
**Cpr1** homolog in *Aspergillus niger* was transcriptionally up-regulated by 2-fold during chemically-induced secretion stress. Moreover, it was post-transcriptionally up-regulated by 7-fold in response to the stress, more than any other gene in their study. The mechanism of this is unknown, but could involve features of the 3'UTR, which has been implicated in regulating translation efficiency and mRNA targeting (reviewed by Grzybowska et al. 2001). Post-transcriptional regulation could also involve alternate splicing of the transcript, leading to production of different proteins or alterations in transcript stability and translation efficiency during different phases of development. In humans, the SPC18 transcript is subject to alternative splicing that results in production of six different protein isoforms (Oh et al. 2005). The roles of these isoforms, and particularly whether they have different functions in protein processing, have not been investigated in the literature, to my knowledge.

Yeast Spc3p contains an N-glycosylation post-translational modification (PTM) *in vivo* (Meyer and Hartmann 1997). However, mutation of the glycosylated residues in yeast had no discernable effects on viability in normal culture conditions (VanValkenburgh et al. 1999). The mammalian homolog of CPR1, SPC22/23 occurs *in vivo* as two versions with different sedimentation coefficients, due to differences in glycosylation (Shelness et al., 1994). Protein glycosylation is associated with functions in protein stability, localization, and complex formation (Freeman, 2000; Roth et al., 2012). During pathogenesis, it is possible that the post-translational modifications lead to conformational shifts in the structure and function of CPR1 in *C. graminicola*.

In this chapter, my goal was to characterize the precise nature of the MT and WT *Cpr1* gene transcripts *in vitro* and *in planta*. I tested three predictions related to the hypothesis that alternate splicing plays a role in CPR1 function and the MT phenotype: i) alternative splicing of the *Cpr1* transcripts occurs during development *in planta* versus *in culture* in WT, and the MT fails to undergo this splicing normally; ii) alternate splicing occurs in the MT but not the WT *in planta*; or iii) alternate splicing does not occur in either strain *in planta*: MT and WT transcripts differ only in the 3'UTR sequence. I also conducted a bioinformatic and comparative transcriptomics study of the conserved SPC and secretory pathway genes in *C. graminicola*, as well as a comparative analysis of the SPC
proteins of *C. graminicola* which included identification of conserved PTM and catalytic residues.

This chapter contains a detailed description of the methods that were used to produce and analyze the *C. graminicola in planta* transcriptome data that I have used throughout my dissertation. As is typical for such studies, the generation of these data was a collaborative effort. Some of the methods and an earlier analysis of the data have been published in O'Connell et al., 2012. I have tried to make it clear in the description that follows what I did and what my collaborators did, and also what has been published previously.
2.3 Material and Methods

2.3.1 Defining the putative secretory pathway in *C. graminicola*

The hypothetical secretory pathway of *C. graminicola* was reconstructed based on similarities to the secretory pathway genes from yeast (Delic et al., 2013; Kienle et al., 2009; Schekman and Rothman, 2002) and filamentous fungal pathogens (Giraldo et al. 2013; Yi et al. 2009; Petre and Kamoun 2014). To identify homologs of these genes in the *Colletotrichum* species, protein sequences obtained from the Saccharomyces Genome Database (SGD) (www.yeastgenome.org) or the NCBI database (for sequences from other filamentous fungi) were subjected to standalone BLASTP with an e-value cutoff of 1e-5. Genes were considered to be orthologs only when the two proteins were reciprocal best hits (RBH) (Nikolaou et al., 2009; Wall et al., 2003).

The genomes of *C. graminicola* and of *C. higginsianum*, published in O’Connell et al. (2012), were downloaded from the *Colletotrichum* Comparative Sequencing Project (http://www.broadinstitute.org/). The genomes of *C. gloeosporioides* and *C. orbiculare* were published by Gan et al. (2012), and both were downloaded from the NCBI database (*C. gloeosporioides* accession number: PRJNA225509; *C. orbiculare* accession number: PRJNA171217). The genome of *C. sublineola* strain CgSl1 was generated in the University of Kentucky AGTC and is currently housed on our laboratory server. Details about these genome assemblies are presented in Table AII.1 of this dissertation.

2.3.2 Sequence analysis of signal peptidase complex proteins

The protein sequences for the four components of the yeast SPC (Spc1p, Spc2p, Spc3p, and Sec11p) were retrieved from the SGD. Homologous sequences in *C. graminicola*, *Magnaporthe oryzae*, *Medicago truncatula*, *Canis familiaris* and *Gallus gallus* were identified by using BLASTP against the NCBI website platform, using the non-redundant (NR) protein sequences database for each species. Alignments of the sequences were made by using the default parameters of the Muscle platform in Geneious. Alignments were not manually adjusted. Sites for glycosylation, myristoylation, and phosphorylation were predicted by using web prediction tools as follows: NetOGlyc 4.0 Server

2.3.3 Fungal strains

The *C. graminicola* strain M1.001 (aka. M2), isolated from diseased maize (Forgey et al., 1978) was obtained from the late Dr. Robert Hanau (Purdue University, West Lafayette, IN, U.S.A.). The nonpathogenic mutant strain (MT) and its complement (MT-C), described in Thon et al. (2002), were both derived from M1.001 (WT). All isolates were routinely cultured on Potato Dextrose Agar (PDA, Difco Laboratories, Detroit) at 23°C under continuous fluorescent light.

2.3.4 In planta visualization of the endomembrane system and *Cpr1* expression

The endomembrane system of *C. graminicola* was labeled by transforming the WT, MT, and MT-C strains with the plasmid pgpdA_Gla514::sGFPThdel, containing GFP linked to an HDEL membrane anchor driven by a constitutive glucoamylase promoter (Vinck et al. 2005). The plasmid was obtained from Dr. C. Van den Hondel. There was no map so I produced one with a combination of restriction digestions and sequencing (Figure AI.5 in Appendix I of this dissertation). I sequenced part of the promoter, the terminator and the plasmid backbone, as well as the region containing the engineered GFP that has a modification in a serine that improves GFP expression in plants (Chiu et al. 1996).

I used the pSITE vectors (Chakrabarty et al. 2007), as modified by Gong et al. (2015) for Gateway technology (Invitrogen), to produce a construct to investigate expression and localization of CPR1 *in planta*. A sequence of 1747 bp comprising the *Cpr1* ORF and its promoter region was PCR-amplified and introduced upstream of the red fluorescent protein (RFP) ORF (Figure 2.2). I introduced the clones into WT *C. graminicola* by Agrobacterium-mediated
transformation (Flowers and Vaillancourt 2005). I recovered several independent transformants and single-spored them before use.

Living hyphae were visualized in vitro after growing them on sterile glass slides in a thin film of Fries complete medium (30 g sucrose, 5 g ammonium tartrate, 1.0 g ammonium nitrate, 1.0 g potassium phosphate, 0.48 g magnesium sulfate anhydrous, 1.0 g sodium chloride, 0.13 g calcium chloride, 1.0 g yeast extract/liter of H2O). Transformants were also observed in vivo, in leaf sheaths inoculated as described below. Transformants were observed with the Olympus FV1000 (Olympus America Inc., Melville, NY, USA) laser-scanning confocal microscope using 543 nm laser line.

2.3.5 DNA Extraction

Fungal biomass for DNA extraction was produced from 5 X 10⁵ spores inoculated into 500 ml of Potato Dextrose Broth (Difco Laboratories, Detroit, PDB) in a 1 liter Erlenmeyer flask grown for 3 days on a rotary shaker at 200 rpm at 23°C. The mycelial mat was collected by vacuum filtration, and 2 grams of the fresh mycelium was ground in liquid nitrogen, until the consistency of talcum powder. The mycelium was mixed with 4 mls of CTAB extraction buffer (20 mls 1 M Tris pH 7.0; 28 mls 5 M NaCl; 4 mls 500 mM EDTA pH 8; 2 g CTAB; 2 mls mercaptoethanol per 100 mls) and incubated at 65°C for 1 hour. After the samples cooled to room temperature, an equal volume of phenol:chloroform:isoamyl alcohol (PCI|25:24:1) was added and the sample was rolled on the orbital mixer table for 5 min, followed by centrifugation at 6000 rpm for 15 min. The upper aqueous phase was removed to a new tube and the PCI extraction was repeated, followed by an extraction with chloroform. The upper aqueous phase was removed to a new tube and the DNA was precipitated with 1 volume of isopropanol. The DNA was spooled from the isopropanol/aqueous mix interface using a bent glass rod. DNA was rinsed several times in 95% ethanol to remove CTAB and dissolved in 1 ml of Tris-EDTA with 5 μl of RNase A solution (10 mg/ml). The sample was incubated at room temperature in the orbital mixer for 30 minutes. A half volume of autoclaved 7.5 M ammonium acetate was added to denature and precipitate proteins, and incubated at room temperature for 30 min. The sample was
centrifuged in a microfuge at top speed, the aqueous phase transferred to another tube and 2 volumes of cold 95% ethanol was added to precipitate DNA. Samples were centrifuged for 30 min in a microfuge and the pellet was rinsed twice with 70% ethanol. After being air dried, DNA was resuspended with 100 μl of sterile Milli-Q water.

2.3.6 Plant growth and inoculations

Maize leaf sheaths of the maize inbred Mo940, or of the hybrid sweet corn Golden Jubilee (West Coast Seeds, Canada, product #CN361E), were used to produce the RNA for transcriptome and reverse transcription (RT)-polymerase chain reaction (PCR) analysis. The samples for the transcriptome analysis were prepared by me and a former Ph.D. student in our laboratory, Dr. Maria Torres, and the method has already been published in O'Connell et al., 2012. Other samples for this dissertation were prepared by me using the same protocols. Plants were grown in the greenhouse, in plastic Conetainers (Super SC-10 UV stabilized Stuewe & Sons, Inc. Oregon, USA) in a growth medium composed of 60% Pro-Mix BX (Premiere Horticulture, Ltd, Riviere du Loup, PQ, Canada) and 40% sterile topsoil. Plants were watered daily or as needed. Beginning one week after germination, a solution of 150 ppm of 20-10-20 fertilizer (Scotts-Sierra Horticultural Products Co., Marysville, OH) was applied two or three times per week. Leaf sheaths were removed from the V2 leaves at the V3 stage of plant growth and cut into segments approximately 3 inches in length.

Spores from two- to three-week-old PDA cultures were used for inoculations. Two mL of sterile water was added to each plate and the surface was rubbed gently with a sterile plastic minipestle to dislodge the spores. The conidial suspension was filtered through sterile glass wool, and centrifuged for 10 mins at 3000 rpm in a table top centrifuge. The conidial suspension was washed 3 times and the concentration was adjusted to 5 x 10^5 spores/ml. Inoculations were done as described in O’Connell et al. (2012). The leaf sheaths were supported with the midrib sides downward inside Petri dishes lined with wet filter paper. Two 20-μl spore drops were applied to the inside epidermal surface of each sheath, approximately 1 cm apart. The Petri plates were placed inside a clear plastic box lined with moistened germination paper at 23°C under
continuous light. The sheaths were sampled at three stages: ~20 hours after inoculation (hpi) for the pre-penetration appressorial phase (AP); ~36 hpi for the intracellular biotrophic hyphal phase (BT); and ~60 hpi for the necrotrophic hyphal phase (NT). For the MT, only the AP and BT phases were collected, because it doesn’t progress to NT. All sheaths were inspected under the microscope to verify the developmental stage, and trimmed to remove as many of the surrounding uninfected plant cells as possible. For the AP and BT samples, the mesophyll layers below the infected epidermal cells were carefully trimmed away as well, in order to increase the fungal/plant ratio. The NT samples were too fragile to be subjected to this last step. For BT and NT samples, the sheaths were brushed gently with a moistened sterile cotton swab to remove any superficial mycelia. Approximately 8 individual trimmed sheath pieces from each developmental phase were pooled into a microfuge tube, flash-frozen in liquid nitrogen, and maintained at -80°C until RNA extraction. The entire process, from initial observation to flash freezing, took no more than 2 minutes per sheath. Three biological replicates for each of eight treatments (WTAP; WTBT; WTNT; MTAP; MTBT; MT-CAP; MT-CBT; MT-CNT) were prepared for RNA extraction. The strains are WT (wild-type), MT (mutant) and MT-C (complemented) during AP (appressoria), biotrophic (BT) and necrotrophic (NT) stages.

2.3.7 RNA extraction

Total RNA was extracted using the protocol that has been published previously, in O’Connell et al. (2012), following the methods described by Metz et al. (2006) with modifications. Briefly, frozen plant or fungal tissue samples were ground with a plastic minipestle, and RNA was extracted with Trizol reagent (Invitrogen). To increase RNA yields, samples were incubated overnight in isopropanol, followed by 100% ethanol, both maintained at -20°C. Trizol extraction was followed by DNase treatment using the RNeasy Plant Mini Kit (Qiagen). Samples for the transcriptome study were extracted by me and Dr. Maria Torres: the other samples for this dissertation were prepared and extracted by myself using the same method.
2.3.8 Transcriptome: Illumina RNA sequencing and data analysis

Only the WT and MT samples (WTAP; WTBT; WTNT; MTAP; MTBT) were included in the transcriptome study. The preparation and sequencing of the WT samples has been published previously, in O’Connell et al. (2012). Briefly, 300 µg of total RNA from each of three biological replicates of each treatment were submitted to the Texas AgriLife Genomes and Bioinformatics Service Center (Texas A&M System). They prepared libraries by using the TruSeq™ RNA and DNA sample preparation kit (Illumina®). A 7 bp barcode adaptor was added to differentiate the biological replicates, and data were generated from 10 lanes using the Illumina Genome Analyzer IIx. A total of four lanes (i.e. technical replicates) of data were generated for WTAP; one lane each was produced for WTBT and WTNT; and two lanes each were produced for MTAP and MTBT. For the MTBT samples, only two of the three biological replicates produced usable data: the third was discarded (see Appendix I for more details). The technical replicates for each treatment were pooled. Data were processed using the Illumina software CASAVA-1.7.0 for base calling and de-multiplexing, and the final read results were stored as individual files for each sample in FASTQ format. Results from a previous analysis of these data were reported in O’Connell et al. (2012), and have been deposited in Genbank (PRJNA151285).

For the work described in this dissertation, a new analysis of the data was performed by Dr. Noushin Ghaffari of the Texas AgriLife Genomics and Bioinformatics Service Center. This re-analysis was prompted in part by my discovery of an error in the original transcriptome analysis that was most likely due to a sample mix-up at Texas Agrilife (this error is described in more detail in Section AI.1, and Figures AI.1-A1.4, in Appendix I, of this dissertation). Dr. Ghaffari first re-mapped the individual reads onto the C. graminicola M1.001 supercontigs (NCBI Biosample: SAMN02953757) by using the CLC Genomics Workbench (GWB) RNA-Seq analysis tool (http://www.clcbio.com/). I used these mapping data to calculate normalized read counts for each gene for each treatment, following the protocol published in O’Connell et al., 2012. The equation was as follows: normalized read count for gene X in treatment Y = (mapped read count for gene X in treatment Y) / (total read count for treatment Y) * (average read count across all conditions) (Tables 2.1, 2.2). Data for the
WT and MT were normalized separately because the total number of reads was relatively low for the MT in comparison with the WT.

A total of $2.2 \times 10^7$ out of $3.5 \times 10^8$ sequencing reads (6.3%) were mapped by Dr. Ghaffari to the fungal genome (Table 2.3). The remapping resulted in a larger percentage of mapped reads than in the original version reported in O’Connell et al. (2012) (Table 2.3). More than 95% of the annotated *C. graminicola* genes were expressed at some point during infection (defined as surpassing five total reads) (Table Al.1). Eighty-four percent of the genes (10,028/12,006) had sufficient reads to pass the two-stage SRBFF filtering process. After applying the edgeR coefficient parameters (including all the coefficients in the glmLRT function) this number was reduced to 4250 genes that had an FDR of at least 0.5. A total of 3723 genes from this selected gene set were identified as statistically differentially expressed at $\alpha = 0.05$ and with a log2FC bigger than 2, and this set of genes was included in my study (Tables 2.1, 2.2).

To validate the new RNAseq analysis, I utilized some data previously reported by Dr. Maria Torres (2013) who used quantitative real-time reverse transcription polymerase chain reaction (qRT-PCR) to measure the expression levels of fourteen different *C. graminicola* genes. I calculated and plotted the Log2 transcript fold-changes (WTAP:WTBT, WTBT:WTNT, and WTAP:WTNT), measured by both RNAseq and qRT-PCR, across 38 individual differential comparisons to measure the correlation between the gene expression profiles. A linear regression value of $R^2=0.8604$, and a slope of $y=1.036$, indicated that the data were relatively consistent for the two analyses (Figure Al.6 in Appendix I).

I also summarized the occurrence of stress-response genes among sequences that were identified in the microarray experiment described by Tang et al. (2006). This study generated a list of differentially expressed sequence tags from biotrophic hyphae recovered from maize stalks by laser microdissection, which were compared with hyphae growing in culture. The complete data set was never published, but the authors generously shared it with me for my study (Drs. W. Tang and J. Duvick, personal communication). I used the logFC values
provided by the authors, and I considered only values that they considered to be significant. I used BLAST (e-value 1e-5) to match the sequence tags from their dataset with *C. graminicola* genes, which had not yet been annotated when they did their study.

### 2.3.9 Heatmaps

Gene expression patterns were visualized by creating heatmaps using log2 fold changes of genes generated from the transcriptome data. Those data were calculated as described in O’Connell et al. (2012). The expression ratio between the normalized counts of a gene in a developmental stage and the geometrical mean number of normalized reads across all the stages was calculated (Table 2.1, 2.2). The log2FC is derived from this expression ratio and it was used to generate heatmaps of secretory pathway gene expression profiles with the Genesis tool (Sturn et al. 2002).

### 2.3.10 Characterization of the MT *Cpr1* allele

The insertion in the 3’UTR of the *Cpr1* gene in the MT strain had never been characterized in detail. In order to determine the sequence of the MT *Cpr1* allele, I used genomic DNA from the MT and WT as a template for PCR. I designed PCR primers based on my initial hypotheses about the nature of the insertion, including my understanding of the sequence of the REMI plasmid pCB1636 (Figure A1.7 in *Appendix I* of this chapter) which I then progressively modified and refined in accordance with the results of the PCR amplification and sequencing. PCR amplification was carried out using Phusion High-Fidelity DNA polymerase (Thermo Scientific, Waltham, MA, USA) in 20 µl reactions consisting of 0.75 µM of each primer, 0.2 mM dNTP and 0.5 units of Phusion. Thermal cycling was performed as follows: 98° C for 30 sec followed by 30 cycles of amplification at 98° C for 10 sec, annealing temperature of X° C (“X” was set independently for each primer combination) for 30 sec, followed by synthesis at 72° C for 30 sec per kb. All of the primers that were used for this analysis are included in Table Al.1 in *Appendix I* of this dissertation. Some PCR products were cloned in the pGEM®-T Easy Vector from Promega (Madison, WI, USA).
Sequencing of PCR products or clones was done by using the BigDye® Direct Sanger sequencing kit (Life Technologies). All sequencing was performed by the Advanced Genetic Technologies Center (AGTC) at the University of Kentucky by using the 3730 DNA Analyzer (Applied Biosystems). Sequence alignments were done by using Geneious v.6 (Biomatters Ltd.).

2.3.11 Southern Blots

Southern blots were used to further test and refine my hypotheses about the structure of the MT Cpr1 allele. Approximately 1 µg of genomic DNA from the MT strain was digested with restriction enzymes overnight and fragments were separated by electrophoresis on a 1.0% agarose gel for 24 h at 30 V. Restriction enzymes used were EcoRI, Xhol, XbaI, XmnI, Stul, Clal, SphI and EcoRV (New England Biolabs). Genomic DNA fragments were transferred from the gel to a nylon membrane (GE Healthcare Life Sciences) by using an electroblotter (Idea Scientific, Minneapolis). Probes for Cpr1, ampicillin (Amp) and the hygromycin phosphotransferase selectable marker gene (Hyg) were produced by PCR amplification with Taq DNA Polymerase (Life Technologies) and the MT DNA as a template, using the primer pairs CPR1intF3 and CPR1intR3 for Cpr1, 5NEWF4 and 5NEWR4 for Amp, and 2NEWF3 and 2NEWR3 for Hyg (Table AI.1 in Appendix I). Probes were purified by using the Qiaquick PCR purification kit (Qiagen), and labeled by using the DNA Polymerase I Large (Klenow) Fragment kit (Promega) for random primer 32P-labeling. A Typhoon PhosphorImager (GE Healthcare) was used for imaging.

2.3.12 Transcriptome analysis to characterize transcripts of Cpr1 and other SPC genes in planta

Reads obtained from the RNAseq data were used to identify alternatively spliced regions in the genes encoding the four proteins that were predicted to comprise the C. graminicola SPC. RNAseq reads were mapped against the genome of C. graminicola using TopHat version 1.3.1 and default settings. The percentages of fungal reads that aligned to the supercontigs for each condition are included in Table 2.3. The alignments and the splice junctions list were visualized by using the Integrated Genome Browser (http://bioviz.org/igb/index.html ©UNC Charlotte). To identify each of the 4
genes, I used *C. graminicola* supercontigs as my reference sequences and I focused my analysis on the regions surrounding and including the positions of the four SPC genes. Supercontig sequences and information about the location of each gene are available from the Broad Institute *Colletotrichum* Database site (http://www.broadinstitute.org/annotation/genome/colletotrichum_group).

### 2.3.13 Reverse transcription (RT)-PCR

I used RNA extracted from leaf sheaths and from fungal cultures to characterize the *Cpr1* transcripts from WT, MT and MT-C strains by RT-PCR. RT-PCR was performed according to the protocol of Venard et al. (2008), with a few modifications. For each reaction, 2 µg of RNA diluted in 10 µl of water was incubated with 1 µl of oligodT primer for 15 mins at 65°C. Four µl of 5x RT buffer, 2 µl of 0.1 M DTT, 1 µl of 10 mM dNTP, 40 units of RNaseout (Invitrogen), and 1 µl (200 units) of Superscript II (Invitrogen) were added and incubated at 42°C for 1 h, followed by denaturation at 65°C for 15 min. One µl of cDNA was used for each RT-PCR reaction. Two different primer combinations were used for amplification of sequences spanning the *Cpr1* intron: CPR1F and CPR1R; and CPR1F2 and CPR1R2 (Table AI.1 in Appendix I). Some PCR products were cloned in the pGEM®-T Easy Vector from Promega (Madison, WI, USA). Sequencing of clones was done by using the BigDye® Direct Sanger sequencing kit (Life Technologies). All sequencing was performed by the Advanced Genetic Technologies Center (AGTC) at the University of Kentucky by using the 3730 DNA Analyzer (Applied Biosystems). Sequence alignments were done by using Geneious v.6 (Biomatters Ltd.).

To characterize the poly(A) tails, RNA extracted from WT and MT strains grown in culture was used to produce cDNA. The RT-PCR reaction was performed as above, but with a different oligodT primer, which was designed based on the protocol in Ma and Hunt (2015) with modifications. The oligodTPolyA primer with degenerated bases included a specific sequence that, after the production of cDNA, can be used as a binding site for the reverse primer. The gene of interest, *Cpr1*, is amplified by using a specific forward primer designed close to the stop codon of the gene (Table AI.1 in Appendix I). The cDNA was used in a PCR with Phusion polymerase to amplify the poly(A) tail regions of the *Cpr1*
transcripts. The primers used for the poly(A) tail amplification were 2NEWF2 and TailR2EB (Table Al.1 in Appendix I). Sequencing was performed by the AGTC as described above.
2.4 Results

2.4.1 Genomics and comparative transcriptomics of the proposed secretory pathway in *C. graminicola*

The genes encoding proteins that comprise the putative secretory pathway of *C. graminicola* are listed in Table 2.4.

Two of the genes, *SEC61* and *UFE1*, appear to be absent from one of five *Colletotrichum* species, *C. gloeosporioides* (Figure 2.3). When I used BLASTP against the NR database of NCBI, I found a match to *SEC61*, but the gene did not appear to be complete. This suggests that the rest of the gene was not sequenced, or that there is a mistake in the annotation. I also found a match to *UFE1* in a different isolate of *C. gloeosporioides* in the NCBI database. Thus, this gene may be missing only from the strain I downloaded, or else it wasn’t sequenced or annotated in that strain. The latter seems most probable since the quality of that sequence annotation is not particularly high (see Table AII.1 in Appendix II).

The proposed secretory pathway of *C. graminicola* is presented diagrammatically in Figure 2.4, together with the normalized read counts for each gene across the three WT developmental phases and the two MT phases, obtained from the RNAseq data.

The secretory pathway begins with proteins that are involved in processing and translocation of the pre-proteins across the ER membrane and into the ER lumen. These include the SPC, the translocon pore, and various chaperones including BIP/Kar2, which stabilize the proteins and assist with folding in the lumen (Gething 1999; Yi et al. 2009; Delic et al. 2013). The translocon pore allows the entrance of nascent proteins being produced by the ribosome into the ER. BIP is located at the translocon pore and binds to the polypeptide as it enters the ER (Seppä and Makarow 2005). Chaperones bind to proteins to stabilize them, facilitating proper folding (Vitale and Denecke 1999). The SPC removes the signal peptide and the protein is released into the ER lumen. Proteins exit the ER by the COPII pathway, also called anterograde transport, a coated vesicle protein transport path to the Golgi (Schekman and Rothman 2002). The formation of transport vesicles starts with Sar1 that, when catalyzed
by Sec12, becomes activated and recruits the heterodimeric complex Sec23-Sec24 to initiate vesicle formation (Valkonen 2003). This complex then interacts with another Sec13/Sec31 complex to form a coat around the proteins that will be transported. The fusion of the vesicles to the Golgi membrane is mediated by Sar1, catalyzed by Sec23 (Yoshihisa, Barlowe, and Schekman 1993). In the retrograde COP1 transport pathway from the Golgi to the ER, proteins that have ER-retention signals (e.g. HDEL) will return to the ER after processing, including glycosylation, in the Golgi. The primary protein in this pathway is Arf1, which interacts with the Golgi membranes to retrieve the proteins (Paczkowski and Fromme 2014). This pathway is dependent on two SNARE proteins, Sec20 and Ufe1 (Ballensiefen et al., 1998; Schleip et al., 2001).

Several proteins are important for the export of proteins across the plasma membrane. In Figures 2.3 and 2.4 I have included some of those. Sso1 is a SNARE protein localized in the plasma membrane that will interact with Snc1 to allow the secretory vesicles to fuse to the membrane. Yeast syntaxins Sso1p and Sso2p belong to a family of related membrane proteins that function in vesicular transport (Aalto et al., 1993). In a recent paper by Giraldo et al. (2013) on the secretion of effectors in M. oryzae, the Sso1 homolog has a role in hyphal development and effector secretion, and localizes at the Biotrophic Interface Complex (BIC). Sso1 mutants were reduced in virulence.

The exocyst is a complex of eight proteins that is usually localized to areas of active secretion (Munson and Novick, 2006), represented in Figure 2.4. The exocyst complex transport is independent from the Golgi, and was required for secretion of apoplastic effectors (Giraldo et al., 2013). Mutants in Exo70 and Sec5, two components of the complex, have reduced virulence in M. oryzae (Giraldo et al., 2013).

Proteins shown between the Golgi and the vacuole are important for transport between the two organelles and endosome-endosome fusion. There are several vacuole protein sorting genes (Vps) involved in this pathway. Both Chc1 and Clc1, represented in Figure 2.4, are subunits of the coat protein involved intracellular protein transport. Pep12 and Vps1 are involved in fusion events at the endosome (Bowers and Stevens, 2005). Vps33 and Vps34 are important for transport from Golgi to vacuole, and mutants in yeast exhibit vacuole defects
(Banta et al., 1988). In the human pathogen *C. albicans* the phosphatidylinositol 3-kinase Vps34 is required for virulence (Bruckmann et al., 2000). There are several proteins involved in the transport to endosome and to vacuoles in yeast (Schekman and Rothman 2002). Ypt7, Ypt52 and Ypt53 are some of the regulators of fusion in the vacuole, involved in endocytosis (Arlt et al., 2011).

The transcriptome data suggest that all of the genes included in the putative pathway are expressed *in planta* during all stages, with the BiP chaperone the most highly expressed (Table 2.5). Only two genes in Figure 2.4 appear to be differentially expressed among different treatments. In the WT, the homolog of Arf1 (GLRG_01625) is more highly expressed during the AP stage. The Sec12 homolog (GLRG_10268) is more highly expressed during WTNT, and also during AP in the MT. It should be noted that overall expression levels for this gene are low. In yeast, Sec12p is an ER-membrane associated protein that is necessary for the initiation of COPII vesicle formation during ER to Golgi transport (Barlowe and Schekman, 1993). The LCD microarray data included 33 of the putative secretory pathway genes: only three (two in the exocyst complex and one in the COPII pathway) were differentially expressed in biotrophic hyphae versus *in vitro*, according to the microarray analysis.

### 2.4.2 The SPC proteins in *C. graminicola*

Genes encoding putative homologs of the four SPC proteins from yeast were identified from *C. graminicola* and *M. oryzae* (both Pezizomycotina in the kingdom Mycota) *Medicago truncatula* (plants) *Gallus gallus* (birds) and *Canis familiaris* (mammals) *C. familiaris* had two paralogs of the Sec11p, and *M. truncatula* had two of the Spc3p, but *Gallus gallus* and the two fungal species had just one homolog of each of the yeast genes (Table 2.5).

The Sec11 protein (Figure 2.5) is the most highly conserved. A mutational analysis of this gene (VanValkenburgh et al. 1999) identified amino acids that were essential for viability/function, including S44, G67, H83, D103, and D109 (Figure 2.5, Sec11, asterisks). These sites are conserved in the Sec11 proteins of all six species I looked at. The same group identified glycosylation of D109 of Sec11 (Figure 2.5, Sec11, grey box), and this site is also conserved in all six species.
In yeast Spc3, the transmembrane domain is predicted to be between residues 15-34, and the region in contact with the ER lumen to be between residues 35-184. In the paper by Meyer and Hartmann (1997), two glycosylation sites are predicted. These are not conserved in the Spc3 homologs of the other species, and neither match the glycosylation sites predicted for the proteins from these other species (Figure 2.5, Spc3, grey boxes). Mutations in conserved residues, including glycosylation sites, in the Spc3 gene didn’t cause loss of viability, even though in yeast loss of this gene is lethal (VanValkenburgh et al. 1999; Meyer and Hartmann 1997). Relatively few amino acids of the Spc3 proteins are conserved among all species (Figure 2.5, Spc3). Also, it is interesting to notice that the *C. graminicola* and *M. oryzae* proteins have a 25 amino-acid insertion in the domain that is predicted to extend into the ER lumen, which the other species don’t have. These two species also share two glycosylation sites in this region.

The other two proteins of the SPC have very low levels of similarity across the six species, with few regions in common (Figure 2.5, Spc1 and Spc2).

### 2.4.3 Structure of the mutant and wild type *Cpr1* alleles

The upstream flank of the REMI insertion in the MT had already been rescued and sequenced (NCBI Biosample: SAMN02953757), but the downstream flank was not successfully recovered (Thon et al., 2000). I was able to identify several individual reads in the transcriptome data that spanned the junction between the inserted plasmid DNA and the downstream flanking DNA, as well as confirming the upstream flank as it had been published (Figure 2.6C). There were five overlapping reads that crossed the downstream junction, and five that spanned the upstream junction. The overlapping sequences included intact *EcoRI* sites at both junctions of the insertion.

I used this information, together with the Southern blot data of Dr. Thon (Thon et al., 2002) and the sequence of the REMI plasmid pCB1636 (provided to me by Dr. Jim Sweigard) to develop a hypothesis about the structure of the REMI insertion. I developed primers based on that hypothesis and, using a combination of PCR amplification, cloning, and sequencing, I was able to “walk” through most of inserted DNA in the MT *Cpr1* locus, including the upstream and
downstream flanks, the intact *EcoRI* sites at the upstream and downstream junctions, and into the ORF that is immediately downstream of *Cpr1* (GLRG_04963, a 5′-3′ exonuclease) (Figure 2.6B). In the process of doing this work I discovered that the plasmid sequence provided by Dr. Sweigard was incorrect, and that the hygromycin cassette in the REMI plasmid was actually inverted compared with his version. The correct map and sequence of the plasmid is presented in Figure A1.7, in Appendix I of this dissertation.

I used the same techniques to confirm the structure of the WT allele, which matched the predicted structure (Thon et al., 2002) and the results obtained by Dr. Park by using 3′ and 5′ RACE (Figure 2.6B).

The insertion in the MT 3′UTR consists of two nearly complete copies of the REMI plasmid in an inverse (head-to-head) orientation. I was unable to sequence across the head-to-head connection between the two copies of the REMI plasmid, although I tried numerous approaches to amplify and/or clone this region. I did some Southern blots (SB) with eight different restriction enzymes and three different probes, one corresponding to *Cpr1*, another to *Hyg* and the third to *Amp* (Figure 2.7). My goal was to confirm my map, and estimate the size of the fragment that was still missing from my sequence. The results of the SB showed that the map is accurate, since the bands hybridizing to the *Hyg* probe (Figure 2.7A), the *Cpr1* probe (Figure 2.7B), and the *Amp* probe (Figure 2.7C) matched the predicted numbers and sizes. The results of the SB suggest that there is not an intact *EcoRI* site at the head-to-head junction between the two plasmid copies, but that the two copies are nearly complete. It seems that I am missing approximately 300 bp from my sequencing of the insertion, based on the sequence I do have and the estimated size of the *Xmn1* fragment probed with *Amp*. My inability to clone or to sequence this region could be related to its expected inverse repeat structure.

Based on these genomic sequences, I predicted the transcript sequences of MT and WT *Cpr1* using the FGENESH online gene prediction tool. For the WT sequence, the transcript was identical to the Broad prediction, and to Dr. Park’s prediction based on RACE analysis. The WT transcript is predicted to encode a 229 amino-acid protein. But the transcript predicted for the MT is much longer
(3312 bp), and has the potential to encode a 1103 amino-acid protein (Figure 2.8).

2.4.4 MT and WT Cpr1 transcripts in planta

I mapped individual reads from the transcriptome data to my confirmed sequences of the WT and MT Cpr1 alleles, including the 5' and 3'UTRs (Figure 2.9).

The MT Cpr1 transcript appears to be the same as the WT transcript from the 5'UTR through to the first intron splice junction. However, although the transcripts from the WT matched the predicted intron splice junctions, there were several reads in the MT transcriptome that spanned those junctions, suggesting that read-throughs were occurring in those samples. I mapped reads to the splice junctions of the introns in the other putative SPC genes (GLRG_10877; GLRG_03901; and GLRG_04022). There were a few variants (although no read-throughs) in both WT and MT for all of these, but only the MT Cpr1 transcripts showed evidence of frequent aberrant intron splicing resulting in read-throughs (Figure 2.10). A recent paper confirmed that Cpr1 and the other three SPC genes are expressed in maize leaf blades during infection, and also confirmed the predicted WT intron splicing patterns for all of these SPC transcripts (Schliebner et al., 2014).

Transcripts mapped to the Cpr1 3'UTR sequence matched the predicted transcript and the 3'RACE results in the WT, but the 3'UTR sequence of the MT was much more complex. Reads were found matching sequences from the entire insertion (Figure 2.9). The pattern of reads suggested that the Hyg genes and areas of the pBluescript in the inserted DNA were being expressed in the MT in planta. It is important to point out, though, that it is difficult to interpret results of my transcript mapping to the MT 3'UTR because it was an inverted repetitive sequence and I was unable to differentiate transcripts from the positive versus negative strands.
2.4.6 RT-PCR amplification and sequencing of alternative transcripts from *C. graminicola* mutant strain

I used RT-PCR to amplify the region spanning the *Cpr1* intron from the WT and MT strains *in planta* during all stages of development, as well as from WT appressoria produced *in vitro*. I consistently amplified at least three different products from the MT, whereas the WT produced only one major amplicon of the expected size (Figure 2.11). One of the MT products was the expected size, but there appeared to be less of this amplicon compared with the WT. Some of the amplified fragments were cloned and sequenced. The WT strain produced only one major amplicon, and only one transcript variant was cloned, which corresponded to the predicted sequence with the intron removed (Figure 2.11).

A cloned fragment from the MT that was the same size as the WT amplicon was also the expected transcript with the intron properly removed. One cloned MT fragment was the same size as the genomic band (470 bp), and in this fragment the intron was retained (Figure 2.11A). One cloned fragment from the MT was smaller than the normal transcript size, and it seemed like a larger intron (close to 370 bp in size) had been removed. I was unable to clone or sequence any bands that were larger than the 470 bp amplicon, so I am not sure what those are. The MT strain produced the same variant amplicons *in vitro*, suggesting that it is a feature of the strain and not dependent on the environment (Figure 2.12).

I used GENEIOUS to predict the ORF that would result if the MT transcript with the intron retained was translated. There is an in-frame stop codon at the beginning of the intron, thus intron retention is predicted to result in production of a 140 aa protein consisting only of the CPR1 N-terminal region (Figure 2.13). This shorter protein still encodes the ER transmembrane region, but it lacks the entire ER luminal domain, and 3 out of 4 predicted glycosylation sites.

2.4.7 Analysis of 3'UTR sequences of *C. graminicola* MT and WT

Cloning and sequencing of the 3'UTR from samples of the MT and WT *in planta* revealed that in both the MT and the WT, it seems to occur as several nested versions (Figure 2.14). I was able to sequence three variants of the WT 3'UTR (Figure 2.14: 1, 2, 3), and two of the MT 3'UTR (Figure 2.14: 4,5). None of the
WT variants that I sequenced matched the predicted transcript that was also confirmed by RACE from in vitro samples. One poly(A) for the MT occurs right after the EcoRI site, and another a few base pairs later. Some of my results from the MT matched areas more than 2000 bp beyond the stop codon so it is possible that in some cases, the 3'UTR in the MT is extremely long. I was not able to obtain continuous sequence for these putative 3'UTRs however, so they are not included in my figure.

2.4.8 In planta analysis of the endomembrane system and expression and localization of CPR1

HDEL is an amino acid motif that anchors proteins in the ER membrane. A construct was produced by Dr. C. Von den Hondel that encodes GFP linked to the HDEL anchor, driven by a strong constitutive promoter. I used this construct to transform the WT, MT, and MT-C strains, and I used the transformants to inoculate maize leaf sheaths. Transformation with this construct allowed me to visualize the putative endomembrane system in the living fungi in planta (Figure 2.15). A similar pattern of fluorescence, which was “netlike”, was seen in all three strains, reminiscent of the appearance of the endomembrane network as reported in the literature (Hickey et al., 2004). Because the hyphae were alive I could see dynamic movement in the system, including apparent vesicle transport. There were no obvious differences between the strains in the structure or behavior of the putative endomembrane system, except that the fluorescence seemed to be somewhat more intense in the MT when growing in the plant. I anticipate that these strains could be valuable for future analyses of the movement of proteins through the secretory system of *C. graminicola.*

The CPR1 protein is predicted to be located in the ER membrane. I made a chimeric construct to express the CPR1 protein complexed with RFP, and transformed it into the WT strain. Because the MT strain is already resistant to the selectable marker that was available for the vector system, I did not transform the MT. Examination of the transformants in planta showed very little red fluorescence, indicating very low levels of the chimeric protein in the WT strain. However, faint fluorescence could be seen in some cases that seemed to correspond with large vacuolar structures (Figure 2.16). Unfortunately,
because the expression was so low, it was not possible to tell if the pattern of fluorescence was similar to that in the GFP-labeled endomembrane system. In future it will be good to transform the MT strain, because it is possible that the WT CPR1 protein “outcompetes” the RFP chimeric version in the SPCs of the WT transformants, causing the low fluorescence that was observed.
2.5 Discussion

The ability to secrete proteins is crucial for pathogenicity in fungi. Yet, we know relatively little about fungal secretory pathways outside of a few, mostly non-pathogenic, model systems. A lack of the critical CPR1 signal peptidase protein in *C. graminicola* should be incapacitating, and yet the MT is nearly normal in culture and during the early stages of pathogenicity. It appears to be deficient specifically in the establishment of biotrophy. How can we explain this? The goal of the work I have described in this chapter of my dissertation was to help me to address this question.

In this chapter I characterized the putative secretory pathway of *C. graminicola*, and evaluated its expression *in planta*. *C. graminicola* seems to have homologs for all of the proteins in the canonical secretory pathway, and they seem to be expressed at similar levels across all stages of development *in planta*. My analysis of the laser capture microdissection (LCD) data from Tang et al. (2006) suggested that expression of secretory protein genes is generally similar *in vitro* and *in planta*. There was one gene, Sec23, a COPII subunit that was highly expressed in the biotrophic hyphae when compared to *in vitro* hyphae. In yeast this protein is necessary for vesicle budding and transport from the ER membranes (Schekman and Rothman, 2002). The homolog of Arf1, in COP1, was differentially expressed during WTAP in the RNAseq data. Differential expression of these two genes indicates that vesicle trafficking between the ER and the Golgi is very active during *in planta* infection. Two genes in the exocyst complex, homologs of Sec3 and Sec5, were differentially regulated in the LCD data. Sec3 was more highly expressed *in planta* while Sec5 was more highly expressed *in vitro*. In *M. oryzae*, Giraldo et al. (2013) found that, while apoplastic effectors were secreted by the conventional ER-Golgi secretion pathway, apoplastic effectors appeared to be exported via the exocyst complex.

I detected very few differences in the expression of secretory pathway genes between the WT and the MT, even though the MT is believed to be affected in the function of the signal peptidase, the first step in that pathway. Apparently, even if the function of the pathway is affected by the mutation, the relative expression of the pathway genes, including of Cpr1 itself, is not. We might expect the need for secretion to be greater during necrotrophy when fungus is
producing lots of cell wall degrading enzymes, but if this is so it also isn’t reflected at the transcriptional level, at least not so I can detect it.

Some organisms have multiple copies of Cpr1 and/or other components of the SPC. For example, mammals have more than one paralog encoding the catalytic Sec11 protein, and they also produce additional versions by alternative splicing in some conditions of the transcripts. If C. graminicola also has the potential to produce multiple isoforms of these proteins, and some of them function specifically in planta, this could explain the conditional nature of the MT. I investigated CPR1 and the other SPC proteins in C. graminicola, and I determined that each exists as only a single copy. Furthermore, I found no evidence, either in the literature (Schliebner et al. 2014), or from my own work, that the transcripts for any of the SPC genes, including Cpr1, undergo alternative splicing in the WT in planta to produce additional proteins. Thus, this does not seem to be a likely explanation for the behavior of the MT.

The removal of introns from transcripts is necessary for the production of functional proteins. Splicing usually begin co-transcriptionally and is completed before the poly(A) tail is added, but there are also many genes where splicing happens post-transcriptionally (Brugiolo et al., 2013). Alternative splicing is common, and it provides a mechanism by which the same protein-coding region of the DNA can produce different protein isoforms, potentially with different functions. Additionally, some transcript variants do not encode proteins but play important regulatory roles. Around 95% of the human protein-coding genes appear to undergo alternative splicing (Chen and Manley, 2009). Intron splicing, performed by a protein and RNA complex called the spliceosome, is regulated by both trans-acting proteins and cis-acting regulatory sites in mRNAs. The 3’UTR region, in particular, has an important role in regulation of translation by alternative splicing. The fact that the MT has an altered 3’UTR sequence led me to speculate that splicing of the Cpr1 mRNA could also vary in the MT compared with the WT.

I set out to test three possibilities related to the hypothesis that alternative splicing of the Cpr1 intron played a role in the function of CPR1 and the MT phenotype: i) alternative splicing of Cpr1 occurs during development in planta versus in culture in WT, and the MT fails to undergo this splicing normally; ii)
alternate (aberrant) splicing occurs in the MT but not the WT in planta; or iii) alternate splicing does not occur in either strain in planta: MT and WT transcripts differ only in the 3'UTR sequence. Analysis of the WT suggested that it never underwent alternate splicing in any condition, whether in vitro or in planta. Thus, I reject the first possibility. I did find evidence for alternate splicing of the MT Cpr1 transcript, both in vitro and in planta, thus the third possibility can also be rejected and the second possibility is supported. Furthermore, it appears that the MT undergoes alternative splicing constitutively, under all conditions that I examined, based on the RT-PCR results, not just in planta. Thus, it is unlikely to be related to specific regulatory conditions in planta, but rather to an innate characteristic of the mutant.

I determined that the MT does encode different cpr1 3'UTR sequences than the WT, some of which appear to be shorter, and others much longer. These results are consistent with the Northern blots of M. Thon that showed the presence of multiple transcript species that were both larger and smaller than the WT in vitro. They are also consistent with the results of M. Torres, who showed no statistical difference in the amounts of transcript between different treatments when she used primers at the 5' end of the gene for real-time RT-PCR. Defects in intron splicing or changes in 3'UTRs would not affect the 5' ends of the transcripts. Since the amount of transcript appears similar, based on both the real-time results of Dr. Torres and the RNAseq analysis, it appears that there is no deficiency in the transcriptional activation of the cpr1 gene in the MT.

It is possible that the aberrant transcripts produced by the MT could be unstable and quickly degraded (although the RT-PCR and Northern results seem to argue against this), or they might be targeted incorrectly, and thus fail to be translated. If the aberrant transcripts are translated, they are predicted to produce an alternate version of the protein that lacks the entire C-terminus, including the ER luminal portion. If this variant protein is stable and can be inserted into the ER membrane, this would certainly affect its ability to bind to other proteins in the lumen including SEC11p, which could destabilize the SPC. The yeast SPC3p homolog of CPR1 has been shown to contain an N-glycosylation post-translational modification (PTMs) in vivo. The CPR1 sequence also includes strongly predicted sites for glycosylation. Protein
glycosylation is predominantly associated with stability, localization and complex formation. During biotrophy and stress response, it is possible that post-translational modifications lead to conformational shifts in the structure of CPR1 that are important for its function. The MT protein lacks 2 of the 3 predicted glycosylation sites, and it's function could also be affected by that difference. The MT does make some normal transcript, but apparently in reduced amounts. It is possible that there is enough of the normal CPR1 protein to support growth in vitro, but not during pathogenicity. In the future it will be important to examine the CPR1 protein directly in both the MT and WT strains.

Why is intron splicing altered in the mutant? Intron splicing is regulated by elements that can be located within the open reading frames (ORFs), but more frequently are found in the untranslated regions (UTRs) i.e. the 5'cap and the 3'poly(A) tail (Mignone et al., 2002). These two structures are essential for efficient processing of the mRNA, and their removal or alteration can cause rapid degradation of the molecule.

Although the 5'UTR contains many important cis-acting elements that can affect translation rates, including e.g. secondary structures and alternative start sites, the 3'UTR is regarded as the main factor regulating mRNA stability. This region is subject to a variety of different regulatory mechanisms, including: a) poly(A) tail length and positioning, b) RNA transport and subcellular localization of the transcript c) initiation of translation, by interacting with the 5'UTR (Mazumder et al., 2003), d) presence of cis-elements where proteins can bind and block the ribosome, e) Adenosine-Uridine-rich elements (AREs) can cause translation inhibition and decay and f) microRNAs (miRNAs) can bind to sequences in the 3'UTR and block translation (Hughes 2006).

Polyadenylation involves two steps in mRNA processing; first is the recognition of specific cleavage site, which is followed by the polymerization of the adenosine tail (Lutz and Moreira 2011). There are five cis-acting DNA elements that are involved in polyadenylation, and they have roles in mRNA stability, tissue-specific expression, translation, export and cellular localization, and miRNA targeting. Defects in polyadenylation are implicated in some human diseases, such as cancer (Lutz and Moreira 2011; Paillard and Osborne 2003). Polyadenylation has been widely studied in oncogenes. Some studies show
that cancer cell lines with shorter mRNA isoforms have higher stability and translation rates due to loss of regulatory miRNA binding sites (Mayr and Bartel, 2009). Studies in lymphocytes show that longer 3'UTRs decreased protein translation efficiencies (Sandberg et al., 2008). In fungi the signal for polyadenylation contains an A-rich sequence (usually AAUAA), 13 to 30 nucleotides upstream from the cleavage site. It appears that the change in the 3'UTR of our mutant fungus has caused the creation and removal of cleavage and polyadenylation points (Ozsolak et al. 2010), creating alternative polyadenylated forms. Both the MT and WT Cpr1 transcripts seemed to occur in planta as multiple polyadenylated forms. In yeast it was observed that 72% of the genes had multiple polyadenylation sites (Ozsolak et al. 2010). It is possible that the different UTR sequences play important regulatory roles in the WT, which are no longer functional in the MT with its different UTRs.

The mRNAs are exported from the nucleus to the cytoplasm for protein synthesis. Specific sequence elements like splicing signals, 5'cap and poly(A) tails are important for RNA transport. Different mRNA types form complexes with different RNA-binding proteins to be exported thru the nuclear pore complex (Cullen 2000). That could also be a factor in the MT, as the alterations in the poly(A) tail might significantly affect the transport of the transcripts in the cell.

Proteins that bind to cis-elements in the 3'UTR might be involved directly in fungal pathogenicity. Franceschetti et al. (2011) studied a trans-acting RNA recognition motif (RRM) protein in M. oryzae that, when mutated, caused alterations in the 3'UTR processing of target mRNAs, leading to a lack of virulence and overall defects in fungal development, secondary metabolism, protein secretion and cell wall biosynthesis. Other regulatory elements that bind specifically to 3'UTRs include microRNAs (miRNAs). MicroRNAs are small ~21 noncoding RNAs that regulate gene expression postranscriptionally by binding to cis-elements in the 3'UTR of genes and either cleaving the mRNA or repressing the target gene (Fabian et al., 2010). Kang and collaborators (2013) identified 13 miRNAs candidates and reported that their expression patterns were associated with cellulase production in Trichoderma reesei. The targets for these miRNAs included 3'UTRs of genes involved in transportation, and
enzymatic transcriptional and translational regulation, among others (Kang et al., 2013). It is possible that the changes in the 3'UTR sequence of the MT have resulted in alterations in patterns of gene regulation due to a lack of binding sites for protein or miRNA regulators.

The work in this chapter has led to the development of a new hypothesis, that the alteration in the 3'UTR sequence of the MT C. graminicola Cpr1 gene results in a reduction in the amount of CPR1 protein produced in planta, and that this results in an inability to establish biotrophy. An alternative hypothesis is the change in the 3'UTR leads to aberrant splicing, resulting in production of alternative protein isoforms that are not functional for establishment of biotrophy in planta. These hypotheses need to be tested by direct investigations of the CPR1 protein, something that is planned for the future. My work in this chapter also developed a model for the secretory pathway in C. graminicola, and developed some fungal strains that can be used to visualize the endomembrane system and CPR1 protein expression and localization in the living plant-fungal interaction. These tools will be valuable for future studies of this interesting and important pathogenicity mutant.
Figure 2.1 Illustration of Cpr1 and the signal peptidase complex. A) Representation of Cpr1 (dark green, intron in light green) in the WT and MT strains, with the mutation site 19 bp after the stop codon. Dark blue boxes show the 5’UTR and 3’UTR, characterized by Dr. Eunyoung Park using RACE. Dark arrows show primers used in Torres et al. 2013 and lighter arrows show primers used in Thon et al. 2002. B) Representation of the signal peptidase complex as well as other structures, such as the translocon, involved in pre-protein processing. SEC11 is dark green and the mutated subunit SPC3 is represented in red. Both SPC1 and SPC2 are represented in lighter green. The drawing also depicts a nascent polypeptide chain and how it is guided through the translocon inside the ER to have the signal peptide removed by the SPC.
Figure 2.2 Illustration of the pFPL Gateway vector construct. It was used to create protein fusions between *C. graminicola Cpr1* promoter region and open reading frame to the red fluorescent protein reporter gene. Modified from figure kindly provided by Dr. Mark Farman.
<table>
<thead>
<tr>
<th></th>
<th>Yeast</th>
<th>Cgram</th>
<th>Csub</th>
<th>Chig</th>
<th>Corb</th>
<th>Cglo</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIP</td>
<td>100</td>
<td>76</td>
<td>73</td>
<td>76</td>
<td>75</td>
<td>72</td>
</tr>
<tr>
<td>SAR1</td>
<td>100</td>
<td>75</td>
<td>73</td>
<td>71</td>
<td>74</td>
<td>75</td>
</tr>
<tr>
<td>YPT7</td>
<td>100</td>
<td>67</td>
<td>59</td>
<td>67</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td>SEC61</td>
<td>100</td>
<td>63</td>
<td>63</td>
<td>63</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>VPS1</td>
<td>100</td>
<td>60</td>
<td>59</td>
<td>61</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>SNC1</td>
<td>100</td>
<td>60</td>
<td>59</td>
<td>60</td>
<td>59</td>
<td>60</td>
</tr>
<tr>
<td>YPT52</td>
<td>100</td>
<td>59</td>
<td>53</td>
<td>52</td>
<td>57</td>
<td>58</td>
</tr>
<tr>
<td>ARF1</td>
<td>100</td>
<td>59</td>
<td>57</td>
<td>45</td>
<td>58</td>
<td>77</td>
</tr>
<tr>
<td>SEC13</td>
<td>100</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>57</td>
<td>58</td>
</tr>
<tr>
<td>CHC1</td>
<td>100</td>
<td>55</td>
<td>55</td>
<td>60</td>
<td>55</td>
<td>56</td>
</tr>
<tr>
<td>SBH1</td>
<td>100</td>
<td>55</td>
<td>55</td>
<td>55</td>
<td>57</td>
<td>55</td>
</tr>
<tr>
<td>SEC11</td>
<td>100</td>
<td>53</td>
<td>47</td>
<td>54</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>SEC23</td>
<td>100</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>SEC24</td>
<td>100</td>
<td>47</td>
<td>47</td>
<td>46</td>
<td>46</td>
<td>47</td>
</tr>
<tr>
<td>YPT53</td>
<td>100</td>
<td>46</td>
<td>44</td>
<td>47</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>SSS1</td>
<td>100</td>
<td>46</td>
<td>48</td>
<td>47</td>
<td>46</td>
<td>46</td>
</tr>
<tr>
<td>VPS34</td>
<td>100</td>
<td>39</td>
<td>39</td>
<td>41</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td>SED5</td>
<td>100</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>38</td>
<td>39</td>
</tr>
<tr>
<td>BET1</td>
<td>100</td>
<td>38</td>
<td>37</td>
<td>38</td>
<td>34</td>
<td>41</td>
</tr>
<tr>
<td>UFE1</td>
<td>100</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>LHS1</td>
<td>100</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>33</td>
<td>32</td>
</tr>
<tr>
<td>SEC22</td>
<td>100</td>
<td>34</td>
<td>31</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>SEC31</td>
<td>100</td>
<td>33</td>
<td>32</td>
<td>30</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>SPC3</td>
<td>100</td>
<td>32</td>
<td>32</td>
<td>35</td>
<td>29</td>
<td>33</td>
</tr>
<tr>
<td>SSO1</td>
<td>100</td>
<td>32</td>
<td>31</td>
<td>32</td>
<td>31</td>
<td>32</td>
</tr>
<tr>
<td>SPC2</td>
<td>100</td>
<td>32</td>
<td>33</td>
<td>34</td>
<td>32</td>
<td>29</td>
</tr>
<tr>
<td>SEC3</td>
<td>100</td>
<td>31</td>
<td>33</td>
<td>33</td>
<td>34</td>
<td>32</td>
</tr>
<tr>
<td>SEC63</td>
<td>100</td>
<td>30</td>
<td>28</td>
<td>27</td>
<td>31</td>
<td>30</td>
</tr>
<tr>
<td>PEP12</td>
<td>100</td>
<td>29</td>
<td>28</td>
<td>27</td>
<td>29</td>
<td>28</td>
</tr>
<tr>
<td>SEC10</td>
<td>100</td>
<td>27</td>
<td>28</td>
<td>28</td>
<td>25</td>
<td>29</td>
</tr>
<tr>
<td>CLC1</td>
<td>100</td>
<td>27</td>
<td>26</td>
<td>29</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>SPC1</td>
<td>100</td>
<td>24</td>
<td>23</td>
<td>26</td>
<td>30</td>
<td>25</td>
</tr>
<tr>
<td>VPS33</td>
<td>100</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>SEC15</td>
<td>100</td>
<td>24</td>
<td>25</td>
<td>26</td>
<td>23</td>
<td>26</td>
</tr>
<tr>
<td>EXO70</td>
<td>100</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>EXO84</td>
<td>100</td>
<td>23</td>
<td>23</td>
<td>23</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td>SEC20</td>
<td>100</td>
<td>22</td>
<td>22</td>
<td>21</td>
<td>21</td>
<td>19</td>
</tr>
<tr>
<td>SEC8</td>
<td>100</td>
<td>22</td>
<td>22</td>
<td>21</td>
<td>21</td>
<td>22</td>
</tr>
<tr>
<td>SEC6</td>
<td>100</td>
<td>21</td>
<td>21</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>SEC5</td>
<td>100</td>
<td>21</td>
<td>22</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>

Figure 2.3 Degree of conservation of *Colletotrichum* secretory pathway based on yeast proteins. Values represent percent sequence identity using BLASTP. Cgram = *C. graminicola*, Csub = *C. sublineola*, Chig = *C. higginsianum*, Corb = *C. orbiculare* and Cglo = *C. gloeosporioides*. 
Figure 2.4 Model of the *Colletotrichum graminicola* secretory pathway together with normalized reads counts for each gene across growth stages. The stages are, in order: WTAP, WTBT, WTNT, MTAP and MTBT. Reads in red mean the gene was differentially expressed between different stages. Reads with light pink or light yellow in the WTBT mean they were differentially expressed in the laser capture data (LCD): pink means higher *in planta* and yellow means higher *in vitro.*
Figure 2.5 Alignment of amino acid sequences of the four signal peptidase subunits. Boxes mark conserved residues of Sec11 in different eukaryotic species. Catalytic residues identified in yeast by VanValkenburgh et al. 1999 are shown with asterisks. Glycosylation sites from Sec11 (Bohni et al. 1988) and Spc3 (Hellmuth and Hartmann, 1997) are indicated by the grey boxes. Glycosylation sites in C. graminicola Spc3 are indicate by orange boxes and were identified using NetOGlyc 4.0 Server.
Figure 2.6 Map of the MT 3’ downstream region. A) Linearized map of plasmid pCB1636 and EcoRl sites. B) Putative MT 3’ downstream region with two copies of the plasmid in opposite directions. C) RNA seq reads matching the EcoRl junction sites. Red nucleotides show where the stop codon is and blue show the plasmid sequences. EcoRl sites are represented in bold letters. RNAseq reads where found aligning to both sites of the mutation. D) Representation of sequencing data for the MT 3’ downstream, showing that all sequences overlapped. I actually sequenced a lot more PCR products across this region, and every part of the insertion except for the area covered by the yellow box was sequenced at least twice in both directions. The shaded light yellow box shows the junction region between the two plasmids that I was not able to sequence.
Figure 2.7 Southern Blots performed using MT DNA. They were probed with A) Hygromycin gene, B) Cpr1 gene and C) Ampicillin gene. D) Map of the restriction sites for each enzyme used in this study.
Figure 2.8 Transcript prediction for WT and MT $Cpr1$ gene using FGENESH. Figures are drawn to scale. Green boxes represent the WT $Cpr1$ exons. Pink boxes represent the predicted exons based on the sequencing of the downstream region of the MT.
Figure 2.9 Illustration of Cpr1 reads for WT and MT strains. The lanes in blue are from the WT and in green from MT. Below the transcripts there is a representation of the genomic map comprising the 3 genes (GLRG_04963, GLRG_04964 and GLRG_04965) that were used to search for transcripts. The peaks show the number of RNAseq reads for the area and the numbers on the left are an approximation of the number of reads that were mapped.
Figure 2.10 Intron splicing in the four signal peptidase subunit transcripts. Blue bars show the mapped RNAseq reads from appressoria and biotrophic stages to each gene and its adjacent sequences. Green bar below Cpr1 represents gene model, with the light green area being the intron. It is possible to observe several reads across the MT-cpr1 intron region, as well as a longer poly(A).
Figure 2.11 Cpr1 intron pattern of WT and MT strains. DNA control is shown for both strains, followed by *in vitro* appressoria (IVAP), appressoria (AP), biotrophic (BT) and for the WT also necrotrophic (NT). Sequencing results from representative clones illustrated here with WT-Cpr1 always showing correct predicted intron and MT-cpr1 with three different versions sequenced from *in planta* growth: normal intron, intron retention and short reads.
Figure 2.12 *Cpr1* transcript intron pattern of WT, MT and MT-C mycelia grown in Fries medium. Both WT and MT-C shown here only in minimal Fries medium as they had the same pattern in all conditions. MT strain treatments are 1. complete Fries medium, 2. complete Fries medium blended, 3. minimal Fries medium and 4. minimal Fries medium blended.
Figure 2.13 Prediction of the MT cpr1 transcript with the intron retained done by Geneious. There is an in-frame stop codon right at the beginning of the intron. Prediction of the poly(A) site was done as in Thon et al. (2000), by looking for consensus polyadenylation signals. The predicted site in the WT is also marked in the figure, as well as the confirmed site based on RACE results.
Figure 2.14 Sequencing of the 3’UTR of the WT and MT strains. WT includes three variants, MT shows two, different variants. RACE results from Dr. Eunyong Park are shown here for comparison.
Figure 2.15 WT and MT-HDEL tagged isolates. They were visualized in the epifluorescence microscope during *in vitro* growth and when inoculated in the plant. Pictures were taken at 400x magnification and 12 and 24 hpi for *in vitro* and *in planta*, respectively.
Figure 2.16 Expression of CPR1-RFP *in planta*. All pictures were taken at 24 hpi with the confocal microscope at 520V/543 nm. Magnification at 400x.
Chapter 3

Comparative genomics of the *Colletotrichum graminicola* secretome and effectorome

3.1 Overview

The literature suggests that secreted proteins, and particularly a class of highly divergent, small secreted proteins known as effectors, are very important in the establishment of infection by plant pathogenic fungi (Stergiopoulos and de Wit 2009; Cantu et al. 2013; Djamei et al. 2011; Hogenhout et al. 2009; Kamoun 2007; Bozkurt et al. 2012). One hypothesis to explain the behavior of the *C. graminicola* *cpr1* mutant is that it fails to produce secreted effectors that are necessary for the successful establishment of biotrophy. So that this hypothesis can be addressed in the long-term, my goal for this chapter of my dissertation was to identify putative *C. graminicola* effectors that are most likely to have a role in the establishment of biotrophy. I used two criteria. First, I assumed that effectors necessary for biotrophy would be expressed early, during pre-penetration and early biotrophic stages of development (Mosquera et al. 2009; Hacquard et al. 2012), versus late in infection, during necrotrophy, when the host tissues are already dead. Thus, in this chapter I have catalogued the putative secreted effector protein genes (the “effectorome”) of *C. graminicola*, and used the *in planta* transcriptome data to identify effector genes that are expressed early during infection. My second assumption was that effectors that played a role in the establishment of biotrophy in living cells would be lineage-specific, because they would be under strong selective pressure (Win et al. 2007; Valent and Khang 2010; van der Hoorn and Kamoun 2008; Ali et al. 2014). Thus, the other part of my work for this chapter involved a comparative analysis of the *C. graminicola* and *C. sublineola* effectoromes. *C. sublineola* is very closely related to *C. graminicola*, but it fails to establish a successful biotrophic infection in maize. Similarly, *C. graminicola* is unable to infect sorghum, which is the host of *C. sublineola*. If both of my assumptions are accurate, then I would expect there to be some overlap between the list of effectors that are expressed early in *C. graminicola*, and the list of effectors that
are divergent between *C. graminicola* and *C. sublineola*. Thus, in this chapter
I tested the hypothesis that the effector genes that are highly divergent in *C.
graminicola* when compared with *C. sublineola* will also be expressed early
during the infection of maize by *C. graminicola*.

### 3.2 Introduction

#### 3.2.1 Effectors and fungal pathogenicity to plants

There are two levels of host defense that pathogens must overcome when
infecting a plant. The first one is known as pathogen-associated molecular
patterns (PAMPs)-triggered immunity (PTI). PAMPs are typically essential
structural molecules that can’t be easily modified, such as flagellin in bacteria,
or chitin in fungi (de Jonge et al., 2011; Koeck et al., 2011; Zipfel et al., 2004).
PAMPs are recognized by plant recognition receptors present in the plant plasma
membrane. Recognition induces PTI which includes a variety of
generalized defense responses. This basal defense is believed to be the reason
why microbes are unable to infect the majority of potential hosts (Göhre and
Robatzek, 2008). In order to cause disease, the pathogens must secrete
various types of molecules (aka. effectors) that overcome PTI (Chisholm et al.
2006; Okmen and Doehlemann 2014). Recognition of these specific effectors
by the host can lead to a secondary level of resistance known as effector-
triggered immunity (ETI) (Thomma, Nürnberger, and Joosten 2011).

Effectors can be broadly defined as pathogen-secreted proteins that have an
effect on host cells, either by altering host-cell structure or by modulating their
function to facilitate infection (effector-triggered susceptibility, ETS) (Ellis et al.,
2009; van der Hoorn and Kamoun, 2008). It is known that some effectors are
translocated and function in the host cytoplasm, where they can target different
host cell compartments (Djamei et al., 2011; Dou et al., 2008; Kemen et al.,
2005; Khang et al., 2010). Other effectors operate in the plant cell apoplast.
Examples include cell-wall degrading enzymes, necrosis and ethylene-inducing
protein (NEP)-like proteins, and small cysteine-rich secreted proteins such as
LysM (de Jonge et al., 2011; de Jonge and Thomma, 2009). One example of
an effector in which the molecular function is understood is Avr2 from *Cladosporium fulvum*, which inhibits host cysteine proteases in the apoplast. There are effectors that can suppress rapid host cell death, deposition of callose, and accumulation of reactive oxygen species (ROS), which are common defense responses that occur during PTI (S. Chen et al., 2013; Gawehns et al., 2014; Gilroy et al., 2011; Hemetsberger et al., 2012; Mengiste, 2012). Other effectors are targeted to the host nucleus, and seem to interfere with transcription or gene regulation (McLellan et al. 2013; Caillaud et al. 2013). The Cmu1 effector (chorismate mutase) of the biotrophic smut fungus *Ustilago maydis* blocks the salicylic acid (SA) pathway in maize plants (Djamei et al. 2011). The activation of the SA pathway normally results in localized cell death, which could block the growth of a biotrophic pathogen. Some effectors that function in the apoplast seem to be NEP-like proteins, which induce host programmed cell death (PCD) and favor necrotrophic growth of the pathogens (Gijzen and Nürnberger, 2006; Kleemann et al., 2012). Some effectors are thought to help the fungus avoid triggering PTI. For example, some mask fungal chitin and thus prevent the induction of chitin-triggered host defenses (de Jonge, Bolton, and Thomma 2011). In general, it is difficult to establish exactly how most effectors facilitate fungal pathogenicity. Effector mutants often don’t have an obvious phenotype, probably due to functional redundancy (Birch et al., 2008; Lawrence et al., 2010; Mosquera et al., 2009). Localization experiments using fluorescent proteins have been done, but the location of the protein doesn’t reveal the specific function of the fungal secreted proteins in the plant (Khang et al. 2010; Kleemann et al. 2012).

3.2.2 The role of effectors in *Colletotrichum*

Both *C. graminicola* and *C. sublineola* are hemibiotrophs, which means that they infect initially as biotrophs and then switch to necrotrophic development. True biotrophs (e.g. rusts, smuts, and powdery and downy mildews) reprogram living host cells by producing small secreted protein (SSP) effectors that suppress PTI and host cell death (Doehlemann et al. 2008; Eichmann et al. 2004; Niks and Marcel 2009, Stergiopoulos and de Wit 2009). Some necrotrophs take advantage of plant defense responses to enhance
pathogenicity, and induce PCD by secreting phytotoxic host-specific (HST) effectors (Amselem et al., 2011; Govrin and Levine, 2002) or toxic metabolites (Navarre and Wolpert 1999). It is suggested that hemibiotrophic *Colletotrichum* fungi first suppress, and then later induce, host PCD (Gan et al., 2012; Kleemann et al., 2012; O’Connell et al., 2012; Stephenson et al., 2000; Yoshino et al., 2012). Arrays of small secreted protein (SSP) effectors that are presumably involved in the initial step, suppression of PCD, are produced by the appressoria and the primary biotrophic hyphae of *Colletotrichum* (Gan et al., 2012; Kleemann et al., 2012; O’Connell et al., 2012).

3.2.3 Identification of putative effectors

Putative effector protein genes can be identified from genome data by using a bioinformatics approach. Effectors in fungi are usually classified as small secreted proteins (SSP), and sometimes more specifically as cysteine-rich SSP (Doehlemann et al. 2009; Ellis et al. 2009). The primary characteristic for bioinformatic identification of an effector is that the protein has an N-terminal sequence that targets it for processing and secretion. Effector proteins are usually described as small, but sources have defined “small” differently, ranging from < 400 amino acids (Bowen et al. 2009) to < 100 amino acids (Kleemann et al. 2012). Some families of oomycete effectors have been identified by the presence of conserved amino acid motifs. Known oomycete effectors contain an RXLR (Arg-X-Leu-Arg) motif, sometimes followed by the dEER (Asp-Glu-Glu-Arg) motif, near the N-terminus of the proteins. The RXLR motif is proposed to be involved in translocation into the host cell (Dou et al., 2008; Whisson et al., 2007; Yaeno et al., 2011). Recently it was found that some oomycetes also contain effectors with additional W, Y and L motifs (WYL) in the C-terminus of the protein (Win et al. 2012; Dou et al. 2008). The WYL confers hydrophobicity to the molecule, but its function in effector activity (if any) is not yet known. Oomycetes also have a second class of effectors, called crinkle effectors, with a conserved motif LxLFLAK that mediates transport into the host. The crinkle effectors were demonstrated to accumulate in the host nucleus, and some of them are required for virulence (Schornack et al. 2010). Some fungal effectors from rusts and powdery mildews have a Y/F/WxC tripeptide motif in
the N-terminal (Rafiqi et al. 2012). This motif appears to be necessary for translocation of the protein from the pathogen to the plant, but not for activity inside the host cell. The Y/F/WxC motif is not universal in fungal effectors, and no other motifs have been confirmed that would facilitate their bioinformatic identification or classification.

Because effectors are assumed to be subject to diversifying selection (Kaschani et al., 2010; Koeck et al., 2011; Maor and Shirasu, 2005; Sperschneider et al., 2014) another common definition of an effector is a sequence that shows no similarity with other sequences in the genetic databases (O’Connell et al., 2012). Because new sequences are added to the databases daily, this definition is something of a moving target. Evaluation of presence/absence polymorphisms or signatures of diversifying selection in proteins in closely related species or strains is another way to identify potential effectors (Rech et al., 2014). The Pathogen Host Interaction database (PHI) catalogues effectors in a variety of pathogenic microbes (Baldwin et al. 2006; Urban et al. 2014). It is possible to use this database to identify conserved or novel members of shared effector families. Another characteristic that can be taken in consideration, if transcriptome data is available, is the timing of gene expression. Effectors that function in pathogen establishment and suppression of PTI during biotrophy would be expected to be expressed early during the infection (Ipcho et al. 2012; Cantu et al. 2013; Duplessis et al. 2011). Furthermore, it appears that many fungal effectors are induced by plant signals and are expressed only in planta, so this can be an additional clue (Marshall et al. 2011).

3.2.4 Use of transcriptome data to identify candidate effectors

There have been several reports on the use of transcriptome analysis for fungal effector identification. In the M. oryzae/rice pathosystem, transcriptome data was used to identify 59 candidate effectors, including the biotrophic-associated proteins (BAS1-4) that accumulate in the biotrophic interfacial complex (aka BIC) (Mathioni et al., 2011; Mosquera et al., 2009). BAS1 was shown to diffuse out ahead of the hyphae into neighboring host cells (Khang et al. 2010). Transcriptome data also helped to identify 437 candidate effectors expressed
in haustoria of the wheat stripe rust *Puccinia striiformis* f.sp. *tritici* (Garnica et al., 2013) and 725 in haustoria and infected leaf tissues of the flax rust *Melampsora lini* (Nemri et al. 2014). Much less is known about effectors in necrotrophic fungi. Transcriptome data generated from tissues infected by the white rot fungus *Sclerotinia sclerotiorum* (Guyon et al., 2014) led to the recognition of 78 candidate effectors expressed *in planta* that might function in blocking host immune responses and inducing PCD. Transcriptome analyses have also been used to characterize oomycete effectors. Thus, Stassen and co-authors (2012) identified 78 RXLR candidate effectors, and a cluster of genes encoding other putative secreted proteins, that were expressed *in planta* by the lettuce downy mildew pathogen *Bremia lactucae*.

The transcriptome data provide a “snapshot” of gene expression at a given time point in the infection. However, transcript levels are not necessarily correlated with protein levels (e.g. Müller et al. 2012). Post-transcriptional regulation, translational controls, and protein stability also significantly affect protein levels. Nonetheless, the transcriptome study is a valuable and necessary first step in designing and interpreting proteomics studies and in designing experiments to functionally characterize individual candidate proteins.

### 3.2.5 The role of effectors in non-host resistance

Effectors can have a positive or negative effect on the disease outcome, from the pathogen’s perspective, depending on the host genotype. Effectors that function in ETI and ETS include a vast array of SSPs in biotrophic and hemibiotrophic pathogens (Giraldo and Valent, 2013), and HSTs in necrotrophs (van der Does and Rep, 2007; Vleeshouwers and Oliver, 2014). ETS is critical early in the interaction, when the pathogen is establishing itself in the host cell, and effector expression usually peaks during early infection (Vleeshouwers and Oliver, 2014). Transcription of effectors is typically induced in response to unknown host signals (de Jonge et al., 2011; Stergiopoulos and de Wit, 2009).

Evidence suggests that inducible non-host resistance in many agriculturally-important pathosystems, particularly in closely related hosts, is actually due to ETI versus PTI: the latter operates more frequently in more distantly related plants. In these cases all members of the non-host plant species contain the
same R gene(s), while all members of the nonpathogenic microbial species contain the corresponding avr gene(s) (Schulze-Lefert and Panstruga, 2011; Tosa, 1992). For example, M. oryzae is divided into a large number of host specific forma speciales (f.sp.). Laboratory crosses have demonstrated that a small number of genes control host specificity among these f.sp. (Chuma et al., 2011; Murakami et al., 2000; Nga et al., 2009; Takabayashi et al., 2002; Tosa, Tamba et al., 2006; Valent et al., 1991). Some of the genes have been cloned, including members of the PWL gene family that controls pathogenicity to weeping lovegrass. They encode highly divergent SSP that meet the definition of effectors/avr gene products (Kang et al., 1995). Crosses have demonstrated gene-for-gene regulation of non-host resistance in both M. oryzae and E. graminis (Matsumura and Tosa, 1995; Takabayashi et al., 2002; Tosa et al., 2006). Comparative genomics has revealed that the majority of differences among host-adapted species and f.sp. in several different biotrophic and hemibiotrophic fungal and oomycete pathosystems are in effector proteins, encoded by rapidly evolving genes (Raffaele et al., 2010; Maryam Rafiqi et al., 2012; Spanu et al., 2010). Work with bacteria showed that type three effectors are involved in host specificity to plants (Hajri et al., 2009; Lindeberg et al., 2009). Effectors in rust have been identified as potentially important for host specificity when compared to closely related rust species (Nemri et al. 2014; Dong et al. 2014). A recent paper by Lee et al. (2014) suggests that Phytophthora infestans effectors might contribute to nonhost resistance. It is proposed that host specificities are determined in many agriculturally-significant pathosystems by repertoires of microbial effectors that have undergone diversifying selection during adaptive co-evolution (Win et al. 2007).

Recent studies that have applied comparative genomics to closely related pathogens have shown that non syntenic regions are enriched with genes encoding polymorphic secreted effector proteins, some of which were shown to be involved in host specificity (P. J. G. M. de Wit et al. 2012; Dong et al. 2014; Nemri et al. 2014; Cantu et al. 2013; Spanu et al. 2010). Work comparing two closely related species of corn smut fungi, Ustilago maydis and Sporisorium reilianum, revealed clusters encoding groups of divergent secreted proteins in the two species, and when the divergent proteins of the clusters were deleted,
some had a negative effect on *U. maydis* virulence in maize (Schirawski et al. 2010). More recent work in *U. maydis* (Brefort et al. 2014) described a cluster containing genes encoding 24 secreted effectors, 12 of which were unique when compared with *S. reilianum* (evalue 1e-10). When these were deleted there was a major negative impact on tumor formation, although the pathogen could still complete its life cycle in the plant. In other work the host ranges of two different species of *Phytophthora* were determined by a single amino acid polymorphism in a protease inhibitor protein (EPIC1) (Dong et al. 2014).

When we inoculate *C. graminicola* in sorghum, it rapidly elicits a visible defense response in the form of accumulation of anthocyanin, a red pigment that is involved in plant defense (Nicholson and Hammerschmidt, 1992). *C. graminicola* is able to germinate and form appressoria on the nonhost, but it cannot establish biotrophic hyphae (Katia Xavier, personal communication). Likewise, when we inoculate *C. sublineola* on maize, the pathogen also cannot establish a biotrophic infection (Torres et al. 2013). However, both *C. graminicola* and *C. sublineola* can complete their life cycles in non-host tissues that have been killed by treatment with herbicides or dry ice (Torres et al. 2013; Katia Xavier, personal communication). This suggests the possibility that effectors produced during the early stages of infection are recognized and trigger non-host resistance (ETI) in the living plant cells.

### 3.2.6 Rationale for identification of *C. graminicola* candidate effectors

Timely secretion of effector proteins in the plant is important for successful infection. For example, when a *M. oryzae* chaperone protein (LHS1) involved in protein folding in the ER was mutated, the fungus was unable to infect rice plants (Yi et al. 2009). The *C. graminicola cpr1* mutant (MT) has an insertion into a gene encoding a putative component of the signal peptidase complex (Thon et al., 2000; Thon et al. 2002). This complex, described in more detail in the previous chapter of this dissertation, is responsible for the first step in the secretory pathway, and regulates entry of proteins into the ER (Fang, Mullins, and Green 1997). Even though *Cpr1* is an essential gene, the MT grows nearly normally in culture (Thon et al., 2002; Venard and Vaillancourt, 2007b). However, when inoculated in maize leaf sheaths, the vast majority (96%) of the
mutant hyphae remain in the first cell and do not establish successful biotrophic infections, or complete their life cycles (Torres et al., 2013). The MT is able to grow on killed maize tissue, and it can also develop apparently normal biotrophic infections if it is inoculated in very close proximity to the WT (Torres et al., 2013). The suggestion was made that one or more diffusible factors produced by the WT strain promotes susceptibility in neighboring cells, and allows the non-pathogenic strain to grow (Torres et al., 2013). These diffusible factors could be effector proteins, some of which have been shown in *M. oryzae* to diffuse out ahead of the infection front (Khang et al. 2010). The MT may be unable to secrete enough of these effectors to establish a compatible interaction with the host.

In order to identify the most likely candidates for *C. graminicola* effectors involved in the establishment of biotrophy, I used a bioinformatics approach and made two assumptions based on the literature. First, I assumed that these effectors would be expressed early, during the pre-penetration and early biotrophic phases of development. And second, I assumed that these effectors would target specific host proteins to suppress PMI and PCD during establishment of biotrophy, and thus would be under selective pressure. If both of my assumptions were true, I would expect some overlap between these two groups of effectors. Thus, for this chapter, I tested the prediction that effectors that are most divergent between *C. graminicola* and *C. sublineola* will also be those that are expressed early during infection. I included *C. higginsianum*, which is proposed to be basal to the clade containing *C. graminicola* and *C. sublineola* (O’Connell et al., 2012; Crouch et al., 2014) in my comparisons. My reasoning was that effectors shared by *C. higginsianum* and by either *C. graminicola* or *C. sublineola*, but not both, might have been lost in one of the lineages due to selection pressure.
3.3 Material and Methods

3.3.1 Fungal strains and genome data used in this study

The *C. graminicola* strain M1.001 was collected in Missouri, USA, in the 1970s (Forgey, Blanco, and Loegering 1978). Strain M5.001 was collected in Brazil in 1989. *C. sublineola* strain CgSl1 was obtained from R. Nicholson of Purdue University, and was collected from grain sorghum in Indiana in the 1970s. The genome assemblies of *C. graminicola* and of *C. higginsianum*, published in O’Connell et al. (2012), were downloaded from the *Colletotrichum* Comparative Sequencing Project (http://www.broadinstitute.org/). The genomes of *C. gloeosporioides* and *C. orbiculare* were published by Gan et al. (2012) and are used in some of my comparisons. Both of these genome assemblies were downloaded from the NCBI database (*C. gloeosporioides* accession number: PRJNA225509; and *C. orbiculare* accession number: PRJNA171217). The genome sequences of M5.001 and CgSl1 were generated by the University of Kentucky (U.K.) Advanced Genetic Technologies Center (AGTC), and they were assembled and annotated by U.K. computational bioinformaticians Dr. Neil Moore, Dr. Jola Jaromczyk, and Dr. Jerzie Jaromczyk. Another *C. sublineola* strain, TX430BB, which was collected from grain sorghum in Texas in the 1980s, was recently sequenced (Baroncelli et al. 2014) and those data were downloaded from NCBI (accession number: PRJNA246670).

3.3.2 DNA extraction protocol

High-molecular weight genomic DNA of strains M1.001, M5.001, and CgSl1 were obtained from cultures grown in 500 ml of liquid Fries Complete Medium (30 g sucrose, 5 g ammonium tartrate, 1.0 g ammonium nitrate, 1.0 g potassium phosphate, 0.48 g magnesium sulfate anhydrous, 1.0 g sodium chloride, 0.13 g calcium chloride, 1.0 g yeast extract/ liter of H2O) inoculated with 1 X 10^5 spores in a 1 liter Erlenmeyer flask on a rotary shaker at 200 rpm for 3 days at 23°C. The mycelial mat was collected by vacuum filtration and 2 grams of the mycelium was ground in liquid nitrogen, until the consistency of talcum powder. The powdered mycelium was mixed with 4 mls of warm CTAB extraction buffer (20 mls 1 M Tris pH 7.0; 28 mls 5 M NaCl; 4 mls 500 mM EDTA pH 8; 2 g CTAB;
2 mls mercaptoethanol perl 100 mls) and incubated at 65°C for 1 hour. After the samples were cooled to room temperature, an equal volume of phenol:chloroform:isoamyl alcohol (PCI/25:24:1) was added and the sample was rolled on the orbital mixer table for 5 min, followed by centrifugation at 6000 rpm for 15 min. The upper aqueous phase was removed to a new tube and the PCI extraction was repeated, followed by an extraction with chloroform. The upper aqueous phase was removed to a new tube and the DNA was precipitated with 1 volume of isopropanol. The DNA was spooled from the isopropanol/aqueous interface using a bent glass rod. The DNA was rinsed several times in 95% ethanol to remove CTAB, and dissolved in 1 ml of Tris-EDTA amended with 5 μl of RNase A solution (10 mg/ml). The sample was incubated at room temperature in the orbital mixer for 30 minutes. A half volume of 7.5 M ammonium acetate was added to denature and precipitate proteins, and incubated at room temperature for 30 minutes. The sample was centrifuged in a microfuge at top speed, the aqueous phase transferred to another tube and 2 volumes of cold 95% ethanol was added to precipitate DNA. Samples were centrifuged for 30 minutes in a microfuge and pellet rinsed twice with 70% ethanol. After being air dried, DNA was resuspended with 100 μl of autoclaved Milli-Q water.

3.3.3 Sequencing and assembly

The genome of M1.001 was sequenced as part of a collaboration with our lab by the Broad Institute of MIT and Harvard (http://www.broadinstitute.org/) to a depth of 9X using a combination of Sanger and 454 sequencing technologies (O'Connell et al., 2012). The genomes of M5.001 and CgSI1 were sequenced by using 454 technology in the U.K. AGTC to 10X and 43X coverage, respectively. Shotgun libraries were prepared according to the "Rapid Library Preparation Method Manual" (Rev 2010). Paired-End 3000 Libraries were prepared according to the "GS FLX Titanium 3kb Span Paired End Library Preparation Method Manual", using a Library Prep Kit, General Library Reagents, and The GS FLX Titanium Paired End Adaptor Set (Roche). Emulsion PCR and enrichment was performed according to the "GS FLX emPCR Method Manual" using the emPCR Kit Reagents (Lib-L) (Roche).
Beads were loaded onto a PicoTiterPlate (70x75) for sequencing with the Sequencing Kit Reagents XLR70 (Roche). Genome assembly was done by Dr. Jola Jaromczyk using Newbler.

3.3.4 Gene annotation

Different annotation methods were used for the different strains. The *C. graminicola* M1.001 genome was annotated by the Broad Institute as described in O’Connell et al. (2012), using a proprietary program called Calhoun that includes a combination of FGENESH (Softberry Inc.), GENEID, and GeneMark, and is trained by using EST data. The *C. graminicola* M5.001 and *C. sublineolea* CgSl1 genomes were annotated by using Maker (http://www.yandell-lab.org/software/maker.html). Maker does *ab initio* prediction, as well as using previous data to train the program to increase the gene prediction confidence. The FGENESH gene prediction program, which is *ab initio* only (Ohm et al. 2010; Salamov and Solovyev 2000), was also used to predict genes in the assemblies of *C. graminicola* and *C. sublineolea*. Gene annotations other than those done by Broad were done by Dr. Neil Moore (Computer Science Department – University of Kentucky).

3.3.5 Transcriptome data

Sample preparation, RNA extraction and sequencing and data manipulation for the transcriptome dataset that I used for my analysis are described in Chapter 2 of this dissertation, and in O’Connell et al. (2012). I also used the laser-capture microarray data from the work of Tang et al. (2006), which was described in more detail in Chapter 2.

3.3.6 Genome synten

Analysis of genome synten between *C. graminicola* and *C. sublineolea* was done by using the Synten Mapping and Analysis Program (Symap) v4.2 and default settings (Soderlund et al., 2006). The 13 chromosomes of *C. graminicola* (O’Connell et al., 2012) were used as a backbone to align and identify syntenic blocks in scaffolds of *C. sublineolea* CgSl1 and *C. higginsianum.
3.3.7 Identification of putative orthologous genes

Two different methods were used to identify putative orthologous genes among the different species. In the first approach I used results generated by the programs Ortho-MCL and Coco-CL (C-CL) (COrrelation COefficient-based CLustering) (Jothi et al., 2006; Li et al., 2003). The program OrthoMCL groups proteins into groups of orthologs and recent paralogs by using a BLAST-based algorithm, and allows for simultaneous analysis across multiple genomes by incorporating the Markov Cluster algorithm (MCL). C-CL is a hierarchical clustering method that does not rely on pairwise sequence comparisons as OrthoMCL does, but utilizes a more global approach that can ‘refine’ the results so that distant paralogs are more accurately excluded from the orthology groups. The species used for OrthoMCL and C-CL comparisons were C. graminicola, C. higginsianum, C. sublineola, M. oryzae, Epichlöe festucae, Fusarium graminearum, F. oxysporum, Trichoderma reesei, Verticillium dahliae and Aspergillus flavus. The OrthoMCL and C-CL analyses were done by N. Moore.

The second approach I used to identify putative orthologous proteins was the Reciprocal Best Hit (RBH) method. This is a common and very simple computational method (Moreno-Hagelslieb and Latimer, 2008; Wall et al., 2003). In RBH, a protein from one organism will be considered to be an ortholog of a protein in another organism if both are the best BLAST hit for one other. A significant weakness of this approach is that it can inaccurately classify distant paralogs as orthologs. RBH is the first step in the OrthoMCL method, but OrthoMCL goes on to apply a weighting protocol to exclude these distant paralogs. However, a major advantage of RBH for my work was that it could be used to characterize all genes. OrthoMCL and C-CL failed to predict orthology groups for some genes, and this was particularly true for the genes that I classified as effectors (see below), many of which don’t appear to have orthologs in the other species that were included in the analysis.

3.3.8 Gene categorization

For both genomes, I used the same programs and parameters to categorize annotated genes.
Genome-wide amino acid similarity analysis of proteins among five different *Colletotrichum* species was performed as described in De Wit et al. (2012). The predicted protein sequences from *C. graminicola* were compared with the other species by BLASTp. Two proteins were considered homologous if they were each other’s best hits. Further, they were only included as homologs if the predicted similarity spanned at least 70% of their lengths, and if difference in length between them was no more than 20%.

To identify members of protein families, I used the Protein Family (Pfam) database (http://pfam.sanger.ac.uk/), with an e-value cutoff of 1e-5 (Punta et al. 2012).

Functional characterization and gene ontology (GO) categories for cellular functions, cellular components, and biological processes, were assigned using the Blast2Go suite (Conesa and Götz, 2008). The GOSSIP function was utilized to determine GO term enrichment in different comparisons (Blüthgen et al. 2005).

To predict secreted proteins, I used Wolf-Psort for fungi, a program that predicts the most likely locations for proteins (www.genscript.com/psort/wolf_psort.html) (Horton et al. 2007). I compared the performance of this program with SignalP (http://www.cbs.dtu.dk/services/SignalP/) by evaluating the ability of both to predict localization of a set of proteins that had been previously functionally characterized as secreted by using the Yeast Sequence Trap Analysis (Krijger et al. 2008).

To identify carbohydrate active enzymes (CAZymes) I used the web resource dbCAN (http://csbl.bmb.uga.edu/dbCAN/annotate.php), an automated CAZyme annotation that is based on the classification scheme of CAZyDB (Cantarel et al. 2009; Yin et al. 2012).

To predict and characterize proteases, I used the MEROPS peptidase database (http://merops.sanger.ac.uk/).

### 3.3.9 Identification of candidate effector proteins

The way that effectors are defined varies in different bioinformatics studies. For example, in the earlier analysis of *C. graminicola* (O’Connell et al., 2012),
effectors were defined as secreted proteins (of any size) that appeared to be unique to *C. graminicola* or to the *Colletotrichum* genus, based on comparisons with the NCBI database at that time. Since 2012 many additional *Colletotrichum* and other fungal genomes have been sequenced, so the list of effectors of *C. graminicola* by that definition has changed and presumably will continue to change. For my work, I defined putative protein effector genes more broadly (and more permanently!) as open reading frames (ORFs) predicted to encode small secreted proteins (SSPs). This was the approach used by Gan et al. (2014) in their analyses of the genomes of *C. orbiculare* and *C. gloeosporioides*. My identification pipeline is shown in Figure 3.1. I utilized the same pipeline to identify putative effectors from *C. graminicola*, *C. sublineola*, and *C. higginsianum*. My first step was to identify predicted secreted proteins by using WolfPsort. My next step was to filter that list to include only proteins that were between 40 amino acids and 300 amino acids in size. Other researchers have had different definitions of SSP, ranging from < 400 amino acids (Bowen et al. 2009) to less than 100 amino acids (Kleemann et al. 2012). I used the same parameters as the ones (Lowe and Howlett 2012) used to identify putative effectors of *Leptosphaeria maculans*. Since there are relatively few functional studies of putative fungal effectors, any decision on what sizes to include is somewhat arbitrary.

Because many fungal effectors have been described as being cysteine-rich (Kleemann et al., 2008; Amaral et al., 2012; Raffaele et al., 2010), I calculated the percentage of cysteines for each SSP, and identified all that had more than 3% cysteine as cysteine-rich (SSP-CR) (Gan et al. 2012).

All the putative protein effector genes were compared with the Pathogen-Host Interaction (PHI database). This database contains “curated molecular and biological information on genes proven to affect the outcome of pathogen-host interactions” from fungi, oomycetes and bacterial pathogens (Baldwin et al. 2006). I used BLAST to identify candidate *Colletotrichum* effector proteins with similarity to any proteins present in the PHI database, with an e-value cutoff of $1 \times 10^{-5}$.

To further characterize the effectors, I used Pfam (http://pfam.sanger.ac.uk/) to identify potential functional motifs. Many fungal effectors lack these
recognizable functional domains, and are typically annotated as “hypothetical proteins” (Sperschneider et al. 2014).

To identify effectors with homologs among the three species, or that appeared to occur in only one of the three, I used a combination of my own RBH analysis, and the OrthoMCL results generated by Dr. Moore. The OrthoMCL results were also used to identify effector homologs that were shared with the other fungal species included in that analysis. In cases where the results of the RBH and OrthoMCL did not agree (there were relatively few of these), I used RBH as the default. Since many of the effectors were not included in the OrthoMCL analysis, this seemed to be the most consistent way to compare across the entire group. The effectors that were found only in C. graminicola (aka “non-conserved” effectors), or that were shared by only two of the three Colletotrichum species, were further evaluated by using BLASTP against the non-redundant protein sequences NCBI database (downloaded in July 2014). Hits with an e-value of below $1 \times 10^{-5}$ were considered to be homologs (Camacho et al. 2009). In a few cases, a different e-value was used and these are explained on a case-by-case basis.

The lists of candidate effector protein sequences for each species were used for standalone BLASTP analysis against five Colletotrichum species (C. graminicola, C. sublineola, C. higginsianum, C. orbiculare, and C. gloeosporioides). The lists were also used to identify putative homologs in M. oryzae 70-15 (MG8) by BLASTP using the the Magnaporthe Comparative Sequencing Project website (http://www.broadinstitute.org/). For the Colletotrichum effectors that had Magnaporthe homologs, a further BLASTP search of the non-redundant protein sequences NCBI database (downloaded in July 2014) was performed to identify those that were only found in Magnaporthe and Colletotrichum and no other genera.

3.3.10 Identification of non-annotated putative effector proteins in C. graminicola

Even though C. graminicola has a very high quality genome assembly, effector proteins can still be difficult to identify, because they are quite small, they often lack functional domains, and they frequently they don’t look like any other genes
in the databases. Annotation programs are trained by using available data sets from previous sequencing projects, and for these reasons they can easily miss effector genes. One solution is to use an \textit{ab initio} annotation method (FGENESH) which does not rely on training with previous datasets. Thus, I used FGENESH to predict additional effector proteins.

3.3.11 Expression analysis of the BAS3 homolog in the WT \textit{in vitro} and \textit{in planta}

3.3.11.1. \textit{In vitro analysis}: The \textit{C. graminicola} WT strain was cultured in 500 ml of Fries complete liquid medium (30 g sucrose, 5 g ammonium tartrate, 1.0 g ammonium nitrate, 1.0 g potassium phosphate, 0.48 g magnesium sulfate anhydrous, 1.0 g sodium chloride, 0.13 g calcium chloride, 1.0 g yeast extract/liter of H$_2$O). Washed spores were added to produce a final concentration of 1x10$^6$ spores/ml, and the culture was incubated at 23°C on a rotary platform shaker at 15 rpm. After 5 days, the cultures were blended and 5 mls of the slurry was added to a new flask containing 50 ml of Fries minimal liquid medium and returned to the shaker. Mycelia were harvested 36 hours later under vacuum filtration and flash frozen in liquid nitrogen, then wrapped in aluminum foil packets and kept at -80°C until RNA extraction.

3.3.11.2. \textit{In planta analysis}: Appressoria of the WT were produced \textit{in vitro} on polystyrene Petri dishes as described by Kleemann et al. (2008), with some modifications. \textit{C. graminicola} spores were collected and washed three times, and 40 ml of a spore suspension at a concentration of 1 x 10$^4$ spores/ml was added to each Petri dish. Twenty hours later, each plate was inspected under the microscope to verify the presence of mature melanized appressoria. Trizol was added and appressoria were broken and scraped from the bottom using a sterile culture spreader. The slurry was recovered from 30 Petri plates in a total of nine ml of TR Izol per replicate.

Infection of maize leaf sheaths by \textit{C. graminicola} and processing of the tissue samples to obtain RNA was done as described in Chapter 2 of this dissertation. The RNA extraction was performed essentially as described in O’Connell et al. (2012), with a few modifications. Frozen mycelia were ground while still contained inside of the foil packet using a pestle. Around 100 mg of the
powdered mycelia was added to a 2 ml Eppendorf tube with 1 µl of Trizol reagent (Invitrogen) for extraction. The cleanup step in the RNeasy Plant Mini Kit (Qiagen) was performed on the supernatant according to the manufacturer’s instructions, including the DNase A treatment.

For the first-strand cDNA synthesis, I used one µg of total RNA and the Superscript II reverse transcriptase kit (Invitrogen) with an oligodT primer. Semi-quantitative RT-PCRs were carried out in 25 µl reactions and consisted of 0.1 µM of each primer, 0.2 mM each dNTP, 0.25 units of Taq DNA Polymerase (Life Technologies) and 1.5 nM MgCl₂. Thermal cycling was performed as follows: 94°C for 3 minutes followed by 30 cycles of amplification at 94°C for 45 s, 60°C for 30 sec and 72°C for 1 min. The primers for BAS3 (EBBAS3F and EBBAS3R) are included in Table AI in Appendix I of this dissertation. Actin (GLRG_03056) was used as an internal control. Sequential dilutions of cDNA were used as template, with the concentrations determined by using amplification of the control gene to normalize across samples, and diluting appropriately so that the control gene was in an exponential range (Choquer et al. 2003).
3.4 Results

3.4.1 Comparison of *Colletotrichum* genomes

3.4.1.1 Genome sequencing and assembly. Genome characteristics of the *Colletotrichum* species used in this chapter are summarized in Table All.1, in Appendix II. Some of these data come from the literature, and some was generated by Dr. Moore and some by me. The total contig length of *C. graminicola*, *C. higginsianum*, and *C. gloeosporioides* was ~50 Mb in each case, whereas *C. sublineola* was a bit larger at ~65 Mb. *C. orbiculare* was the largest, with ~90 Mb. *C. sublineola* has 13,331 predicted genes, a little bit more than the 12,006 predicted in *C. graminicola*. However, when Dr. Moore used MAKER to re-annotate *C. graminicola*, it actually predicted 14,419 genes, more than in *C. sublineola*. *C. higginsianum* is predicted to have more genes than *C. graminicola* and *C. sublineola*, which may be partly due to the fragmented nature of the currently available genome assembly of this species. I estimated that close to 9% of conserved genes were either split or truncated in this *C. higginsianum* assembly (O’Connell et al., 2012). All three genome assemblies contained homologs for most or all of a set of phylogenetically conserved genes (CEGMA) (Parra, Bradnam, and Korf 2007) and a set of conserved fungal genes (Liu et al., 2006) (Table All.1 in Appendix II), suggesting that all of the assemblies are similarly complete.

3.4.1.2 Synteny analysis. Gene order (synteny) is relatively highly conserved between *C. graminicola* and *C. sublineola*. I was able to align 83% of the *C. graminicola* genome assembly with *C. sublineola* scaffolds, and 79% with *C. higginsianum* scaffolds, based on the relative arrangement of conserved genes (Figure 3.2A, Table 3.1). Much of what could not be aligned was comprised of the three *C. graminicola* minichromosomes, which seem to be largely unique to this strain of *C. graminicola* (Rollins 1996). Although the percentages of scaffolds that could be aligned were similar, the number of syntenous genes contained within the aligned sequences was very different. Between *C. graminicola* and *C. sublineola* 85% of the genes were syntenous, but that number dropped to only 50% for *C. higginsianum* genes (Table 3.1). In comparing *C. graminicola* with *C. sublineola*, regions that appear to be inverted,
and regions that appear to lack synteny, could be discerned embedded within the largely co-linear assemblies (Figure 3.2B).

The relative similarity of *C. graminicola* and *C. sublineola* can also be seen in the degree of amino acid identity among predicted proteins (Figure 3.3). Genome-wide amino acid similarity analysis was performed as described in De Wit et al. (2012). Among the proteins shared by *C. graminicola* and *C. sublineola*, 66.4% have more than 81% similarity. Only 44% of the shared proteins of *C. graminicola* and *C. higginsianum*, on the other hand, are that similar. The other two genomes, *C. gloeosporioides* and *C. orbiculare* share even less similarity with *C. graminicola*, with less than 30% of the proteins having more than 81% similarity.

3.4.2 Comparative analysis of *Colletotrichum* proteins

3.4.2.1 Identification of orthologous proteins. Results from OrthoMCL analysis including ten other species of Ascomycete fungi indicated that the three *Colletotrichum* species shared the majority of their proteins, comprising 8799 orthologous groups (Figure 3.4). *C. graminicola* and *C. sublineola* had more groups and more proteins in common than either species shared with *C. higginsianum*. OrthoMCL identified 134 proteins as present in only *C. graminicola*, and 456 that were found only in *C. sublineola* (Figure 3.4). The majority of these non-conserved proteins had no identifiable domains or predicted functions. However, nearly all of them matched sequences in other species, typically identified as hypothetical proteins, in the NCBI databases. Only one protein in *C. graminicola*, and 61 in *C. sublineola*, appeared to be species- or strain-specific orphans.

Approximately 9% of *C. graminicola* genes and 16% of *C. sublineola* genes could not be included in the OrthoMCL analysis. OrthoMCL also failed to characterize 29% of *C. higginsianum* genes. For this reason, I also utilized the Reciprocal BLAST Hits (RBH) approach to analyze orthologous proteins (Wall et al. 2003). With this approach, all proteins can be accounted for. For more than 90% of the proteins, the two methods gave the same result. Results of RBH agreed with results of OrthoMCL analysis in suggesting that a majority of proteins are shared among the three species, and that *C. graminicola* shares
more proteins with *C. sublineola* than either does with *C. higginsianum* (Figure 3.5). The RBH identified more non-conserved proteins than OrthoMCL: 1,164 were found only in *C. graminicola* and not the other two species, and 2,502 were found only in *C. sublineola* (Figure 3.5).

### 3.4.2.2 Protein families analysis

The Protein Family Database (Pfam) (Punta et al. 2012) was used to characterize and compare the predicted proteins from *C. graminicola*, *C. sublineola*, and *C. higginsianum* (Table 3.2). Only 67% of *C. graminicola* proteins, 62% of *C. sublineola* proteins, and 58% of *C. higginsianum* proteins could be categorized into Pfam families. A majority of these families were shared by all three isolates, with relatively few differences in the number of family members across the strains. There were 35 exceptions in which there was at least a three-fold expansion in one or two of the three species (Table 3.2). In cases of apparent gene family expansion, nearly all were expanded in *C. higginsianum* relative to the other two species, or less frequently in *C. higginsianum* and *C. sublineola* relative to *C. graminicola*. One of these differentially expanded families was PF03211, a family of pectate lyases, represented by 14 members in *C. higginsianum* but only four in *C. graminicola* and three in *C. sublineola*. Another family of pectate lyases, PF00544, had twice as many members in *C. higginsianum* than in the other two species. This increased representation of pectin degrading enzymes was confirmed by an analysis of the proteins with the Cazymes database (www.cazy.org) (Figure 3.6). Previously, an expansion of pectate degrading enzyme genes was noted in *C. higginsianum* in comparison to *C. graminicola*, and postulated to relate to differences in dicot versus monocot cell wall structure (O’Connell et al., 2012). Another family expanded in *C. higginsianum* (PF00668) was related to polyketide synthases, which have previously been shown to be more abundant in *C. higginsianum* than in *C. graminicola* (O’Connell et al., 2012). In one case, there was an expansion only in *C. sublineola*, for family PF14529, which is likely to be a retrotransposon-associated gene. In no case was a gene family notably expanded in *C. graminicola* relative to the other two strains.

There were 18 Pfam families that were found only *C. graminicola* and not in the other two species, while 29 were found only in *C. sublineola* (Table 3.2). There were also 121 families that were found in both *C. graminicola* and *C. sublineola*,

86
but not in *C. higginsianum* (Table 3.2). Almost all of these species-specific or clade-specific families contained only a single protein. To my knowledge, none of the species- or clade-specific families have been implicated directly in pathogenicity.

3.4.2.3 BLAST analysis. The *C. graminicola* and *C. sublineola* predicted proteins were further evaluated by using BLAST against the NCBI database (downloaded in July 2014). Of the *C. graminicola* proteins that were conserved among all three *Colletotrichum* species, 64 of them appear to be specific to just these three species, having no significant similarity to other sequences from the database. The majority of the proteins (547 of 560) that were shared only between *C. graminicola* and *C. higginsianum* were present in other organisms (Figure 3.7). Among the proteins shared only by *C. graminicola* and *C. sublineola* most were also present in other organisms (880 of 1018). In contrast, about 50% (550/1164) of the non-conserved proteins in *C. graminicola* that were identified by RBH, and about 40% (1029/2502) of the *C. sublineola* non-conserved proteins, had no matches in the database outside of those species. Among the nonconserved *C. graminicola* proteins with hits, the majority (337/614) were to hypothetical proteins in other species. In *C. sublineola*, 867/1473 of the non-conserved proteins had hits to hypothetical proteins.

3.4.2.4 Characterization of non-conserved proteins. The average size of the predicted non-conserved proteins for both *C. graminicola* and *C. sublineola* was smaller than the average for all proteins [186 amino acids (aa)] vs. 466 aa; and 256 aa vs. 462 aa, respectively). If the analysis was limited to only those proteins with no matches in the database (orphans), the average size was even smaller (109 aa and 180 aa, respectively). Forty percent of the non-conserved proteins of *C. graminicola*, including 37% of orphans, had transcript evidence (defined as a minimum of five normalized reads in at least one sample) (O’Connell et al., 2012) (Table 2.1). We don’t have transcript data for *C. sublineola* so I could not do a similar analysis for those predicted proteins.

More than half of the non-conserved proteins in both *C. graminicola* and *C. sublineola* were predicted to localize to mitochondria or nucleii. About 2/3 of the proteins that had no matches in the database (orphans) in each case were also
predicted to be mitochondrial or nuclear. Somewhat surprisingly to me, only about 10 percent in each case were predicted to be secreted.

A majority of the non-conserved proteins in both species did not have Pfam categories. Among those with Pfam classifications, the largest groups were transporters, cytochrome P450s, carbohydrate-active enzymes (Cazymes), transcription factors, and secondary metabolism enzymes. There was also a large group of proteins in each case that were categorized as heterokaryon incompatibility factors, and a number of proteins that were potentially involved in signaling (i.e. protein kinases and protein phosphatases) and pathogenicity [i.e. proteins with necrosis inducing NPP domains (Gijzen and Nürnberg, 2006; Kleemann et al., 2012), NUDIX domains (Bhadauria et al., 2013), and CFEM domains (Kulkarni et al., 2003)]. About a quarter of the annotated genes in each of the three Colletotrichum species had matches in the PHI database. In the group of non-conserved proteins in C. graminicola, 125 matched the PHI database, and 279 matched the PHI database in C. sublineola.

3.4.3 Comparative genomics of Colletotrichum secretomes and effectoromes

3.4.3.1 The Colletotrichum secretome. To predict the secretome of each species, I used the WolfPsort protein subcellular localization prediction program. In O’Connell et al. (2012) it was stated that this program classified known extracellular proteins better than other programs that are commonly used. I also found that it did a better job than SignalP of predicting localization for a list of C. graminicola proteins that had previously been identified as secreted proteins by using a yeast secretion signal trapping technique (Krieger et al. 2008). WolfPsort predicted that are 1,690 secreted proteins encoded by C. graminicola and 1,891 by C. sublineola, accounting for about 14% of the proteins for each species.

3.4.3.2 The Colletotrichum effectorome. My criteria for calling a protein a putative effector were that it was predicted to be secreted, and that it was between 40 and 300 amino acids in size. I identified 687 small secreted proteins (SSPs) in C. graminicola (5.7% of all proteins); 824 in C. sublineola (6.2% of all proteins); and 1178 in C. higginsianum (7.3% of all proteins). About 40% of the C. graminicola and C. sublineola secreted proteins have fewer than 300 amino
acids. Based on the results from RBH, approximately 400 SSPs are shared among all three *Colletotrichum* species, and there are more proteins shared between *C. graminicola* and *C. sublineola* than *C. graminicola* and *C. higginsianum* (Figure 3.8).

It is interesting that the level of amino acid similarity of homologous secreted proteins is smaller than that of non-secreted proteins (Figure 3.9). If we consider only SSPs versus all secreted proteins, there’s even less amino acid similarity (Figure 3.10). There is consistently less similarity between *C. graminicola* and *C. higginsianum* proteins, versus between *C. graminicola* and *C. sublineola* proteins.

I used the Pfam database to classify the SSPs into functional protein families. Almost half of the SSPs shared by the three species, or shared only by *C. higginsianum* and *C. graminicola*, could be classified. In contrast, only 30% of the SSPs shared between *C. graminicola* and *C. sublineola* could be classified, and that dropped to less than 10% of the non-conserved *C. graminicola* SSPs. The trend is similar in *C. sublineola*.

The *C. graminicola* and *C. sublineola* effectoromes are comprised mainly of hypothetical proteins. Only 36% of all of the *C. graminicola* SSPs, and 31% of the *C. sublineola* SSPs, can be classified by Pfam. The majority of the SSPs have hits on the NCBI database (89% in *C. graminicola*, and 87% in *C. sublineola*) but most in each case hit hypothetical proteins.

I analyzed the cysteine content of the SSPs, as effectors are often described as being cysteine rich (SSP-CR). There are 251 SSP-CR in *C. graminicola*, and 306 in *C. sublineola*. The majority of those are homologous to hypothetical proteins in the NCBI database. Among the SSP-CR, 62 and 53 are characterized by Pfam, and 20 and 13 have similarities to proteins in the PHI database, in *C. graminicola* and *C. sublineola*, respectively.

3.4.3.3 Conserved effector classes in *C. graminicola*. Several classes of fungal pathogenicity effectors described in the literature from other organisms have homologs in *C. graminicola* and *C. sublineola*.

The Common in Fungal Extracellular Membrane (CFEM) proteins have an eight cysteine-containing domain of around 66 amino acids (Kulkarni, Kelkar, and
Some CFEM proteins have important roles in pathogenesis. For example, PTH11 and ACI1 from *M. oryzae* are required for appressorium development (Choi and Dean, 1997; DeZwaan et al., 1999). All three *Colletotrichum* species encode numerous secreted and membrane-bound CFEM proteins. Pfam identified 24 CFEM-domain proteins in *C. graminicola*, and 11 of those are SSP-CRs. *C. sublineola* has 22, 10 of which are SSP-CRs. Homologs of another conserved cysteine-rich secreted effector protein, cerato-platanin, are also found in all three *Colletotrichum* species. Cerato-platanin acts as a toxin in the wilt fungus *Ceratocystis fimbriata* (Pazzagli et al. 2009) and is known as a general fungal elicitor and inducer of PCD.

Chitin-binding proteins contain one or more chitin-binding domains, and they also bind to various complex glycoconjugates (Raikhel et al., 1993). In plants, these proteins are assumed to have a role in host defense, but in fungi they are believed to bind to chitin present in fungal cell walls, thus protecting the pathogen from plant chitinases. The Avr4 protein of *Cladosporium fulvum* is a chitin-binding domain effector that, when mutated, resulted in decreased virulence (van Esse et al., 2007). *C. graminicola* has two genes identified with the chitin binding domains (GLRG_06483 and GLRG_10441). *C. sublineola* and *C. higginsianum* also have two chitin binding domain proteins.

The lysin motifs (LysM) were identified as a class of conserved effectors in pathogenic and nonpathogenic fungi (Kombrink and Thomma 2013). All three *Colletotrichum* genomes contain an expanded family of genes encoding LysM proteins. These genes appear to be highly divergent among the species, and thus to be evolving rapidly (Kleemann et al. 2012). LysM effectors, eg. *Ecp6* from *C. fulvum*, are believed to sequester fungal chitin fragments, thus avoiding host detection (Jonge et al. 2010). The same function was ascribed to *M. oryzae* *Slp1* and *Mycosphaerella graminicola Mg3* LysM effectors (Marshall et al. 2011; Mentlak et al. 2012). In *C. lindemuthianum*, a LysM protein called CIH1 was localized specifically to the surface of biotrophic hyphae by using a monoclonal antibody (Pain et al., 1994; Perfect et al., 1998). All three *Colletotrichum* species have homologs of CIH1. There are six LysM-protein genes in *C. graminicola*, including two SSPs, one of which is the CIH1 homolog.
The Nudix hydrolase, CtNUDIX, was identified in a transcriptome study of the *C. truncatum*-lentil interaction. CtNUDIX was proposed to induce cell death during the switch to necrotrophy (Bhadauria et al., 2011). There are 17 proteins with the NUDIX domain identified by Pfam in *C. graminicola*. Homologs of the Nudix effector are also present in other hemibiotrophic pathogens including *M. oryzae*, and *P. infestans*, but absent in biotrophic and necrotrophic pathogens, prompting the suggestion that it might be important specifically for this lifestyle. CtNudix homologs are also present in *C. higginsianum*, but interestingly, not in *C. sublineola*.

There are several known fungal effector families that induce PCD in plant assays. NIS1 is an effector that is expressed in biotrophic hyphae of *C. orbiculare*, and induces host cell death in the model plant *N. benthamiana* (Yoshino et al. 2012). There is one homolog in *C. graminicola* (GLRG_05338). Homologs of Necrosis Inducing Proteins (NIP), described from biotrophic hyphae of *Fusarium*, are found in all three *Colletotrichum* species. Six genes encoding members of the necrosis- and ethylene- inducing peptide (NEP) 1-like protein family (Gijzen and Nürnberger, 2006) were identified in *C. higginsianum* (Kleemann et al. 2012). However, only three of these homologs actually caused cell death in *N. benthamiana*: the others lacked crucial amino acids and were not able to induce necrosis (Kleemann et al. 2012). Homologs of all but one of the *C. higginsianum* proteins were present *C. graminicola* and *C. sublineola*. It is interesting to note that there are two *C. sublineola* proteins that match ChNLP3 and 3 that match ChNLP5, but *C. graminicola* has only a single homolog for each of these proteins (Table 3.3).

Of the four biotrophy-associated secreted (BAS) proteins described in *M. oryzae* (Mosquera et al. 2009), BAS2 and BAS3 are present in all three *Colletotrichum* species and *C. higginsianum* also has an homolog of BAS4.

CgDN3 is a small secreted protein that is required for the successful establishment of *C. gloeosporioides* on *Stylosanthes guianensis* leaves (Stephenson et al. 2000). Homologs of CgDN3 were found in the genome of *C. higginsianum*, but not in *C. graminicola* or *C. sublineola*. 
3.4.3.4 Orphan genes. Genes encoding secreted proteins involved directly in host-pathogen recognition are frequently highly divergent, because they are subject to rapid adaptive evolution due to selection pressure (Stergiopoulos and de Wit, 2009). Using comparative genomics, I was able to identify lineage-specific proteins ("orphans") that lack similarity to other proteins in the databases. Although these orphan genes potentially have homologs that could be identified in the future as new species are sequenced every day, using a very closely related pathogen in my comparisons should increase the chance that the genes I identified as orphans really are species-specific.

I did a BLAST analysis of all of the proteins from C. graminicola, C. sublineola, and C. higginsianum against the NCBI non-redundant database. It is interesting to notice that out of the 9264 genes shared among the three species, 64 appear to be specific only to those three, including four SSPs. Of the genes that were shared only between C. graminicola and C. higginsianum, all but 13 genes, three of which were SSPs, had hits to other species in the database. Most of the C. graminicola and C. sublineola homologous genes also had hits on NCBI; there were 138 that did not have homologs, out of 1,018 shared genes. Twenty-two of these are SSPs.

It was among the sets of non-conserved genes that I found most of the genes with no homology with any other proteins in the databases. Out of 1,164 non-conserved genes in C. graminicola, 583 (50%) had hits in the NCBI database, with 337 of those being to hypothetical proteins in other species. The 581 remaining genes, including 49 predicted to encode SSPs, appear to be C. graminicola-specific. In C. sublineola, out of the 2,502 non-conserved genes, 1400 (56%) have hits in the NCBI database, 824 of those to hypothetical proteins in other species. Among the 1102 orphan genes, 117 are predicted to encode SSPs.

3.4.3.5. Effector families and clusters. OrthoMCL classified 579 C. graminicola SSPs out of the 687 in groups. Most groups included just one gene copy from each fungus, but a few contained up to 3 paralogs from one or more species. This suggests that there has been relatively little duplication and diversification of these effector proteins. Most of the effectors that were included in OrthoMCL are shared with one or more of the other genera that were incorporated in the
analysis. It was interesting to see that 47 SSPs were found only in *Colletotrichum* and *M. oryzae*, and not in the other Ascomycetes included in the analysis. *M. oryzae* causes rice blast disease and it is a hemibiotroph like *C. graminicola*, with a very similar mode of infection and development in its host. There are two SSPs that are found ONLY in *C. graminicola* and *M. oryzae*, based on BLAST searches of NCBI.

I used the Broad Institute Colletotrichum Database (http://www.broadinstitute.org/annotation/genome/colletotrichum_group/MultiHome.html) to further identify conserved and non-conserved protein families in *C. graminicola* and *C. higginsianum*. Using this tool, I found that 468 *C. graminicola* SSPs are grouped into families in one or both of these species. I used BLAST analysis to determine whether *C. sublineola* also contained members of these families. There were no families that were specific to *C. graminicola*, without members in the other two species. There are seven families in *C. graminicola* that have members only in *C. sublineolum* and not in *C. higginsianum*. An example of one of these families is shown in Figure 3.11A. Most families had members in all three species. For example, I found that the gene GLRG_04750 is part of a family with three paralogs in *C. graminicola* and one in *C. higginsianum*. Using BLAST, I identified four genes in *C. sublineola* that also belong in that family (Figure 3.11B). Some *C. graminicola* effector families also had members in more distantly related species. Thus, there is a family that has members in *C. sublineola* and *M. oryzae* (Figure 3.11C).

I used the program MEME to identify potential protein motifs shared among the SSPs. MEME identified several 10-bp motifs in the *C. graminicola* putative effectors, shared among at least 11 of the proteins (Figure 3.12). Several of the motifs were cysteine rich. I also analyzed the 105 SSPs that are found only in the three species of *Colletotrichum*, but I could not identify any motifs that were consistently shared by these proteins. None of the motifs that I found in the *C. graminicola* SSPs matched any of those that have been reported in the literature (e.g RXLR, dEER, crinkle motifs, etc.). I found no evidence for these motifs in the *C. graminicola* putative effectors.
3.4.4 Effect of effector diversity among isolates

For this dissertation, I had access to genome sequences of two strains of *C. graminicola*, and two of *C. sublineola* (Table AII.1 in Appendix II). Having two strains of each species made it possible for me to compare effector diversity within species.

Only 73 out of 12006 genes (~1%) are not shared between the two *C. graminicola* isolates, and only five of those are SSPs. Of those 73 genes present only in M1.001, 41 of them were found only in *C. graminicola* and not in *C. sublineola* and *C. higginsianum*: 15 were shared with *C. sublineola* only: 2 were shared with *C. higginsianum* only: and 14 were found in all three species. Ninety-nine percent of the genes shared between M1.001 and M5.001 had more than 90% similarity by BLAST. Among the 41 proteins from *C. graminicola* that were not shared with M5.001 or with *C. sublineola* or *C. higginsianum*, 25 had hits to other species using the NCBI non-redundant database, mainly *C. gloeosporioides*, *C. fioriniae* and *Fusarium* species. The remaining 16 proteins appeared to be unique to *C. graminicola* M1.001. Two of these were SSP-CRs.

The *C. sublineola* isolate CgSI1 has 117 genes (less than 1%) that are not shared with TX430BB, and 23 of those are SSPs. Of the 117 genes present only in CgSI1, seven are shared with both *C. graminicola* and *C. higginsianum*: nine are shared with *C. graminicola* only: eight with *C. higginsianum* only: and the majority, 93, were not shared with either species. Using BLAST against the non-redundant nucleotide database from NCBI, I saw that 37 of these non-conserved genes had hits to sequences in other species, mostly *C. gloeosporioides*, *C. orbiculare*, *M. oryzae*, *F. oxysporum*, *Ophiostoma pieceae* and *Podospora anserina*. The remaining 56 genes appear to be unique to CgSI1, including ten SSPs.

3.4.5 Transcriptome analysis: expression of putative effector genes in MT vs WT *C. graminicola* in planta

3.4.5.1. Secreted proteins among the most highly-expressed genes. I generated lists of the 100 most highly expressed genes for each of the treatments (Table AII.2 in Appendix II). Analysis of GO-Terms using Blast2GO
suggested that the majority of the genes in each case were involved in primary metabolism, growth, and signal transduction (Tables AI.3 in Appendix II). About 20% of the most highly expressed genes in each condition were related to stress response. Putative SSP effectors comprised between 9% (WTNT) and 19% (WTAP) of the lists.

When comparing the lists between WTAP and WTBT, there was relatively little overlap, with only 24/100 genes that were shared. In contrast most genes (81/100) were shared between the MTAP and MTBT top 100 lists. A majority of genes (65/100) were also shared between WTBT and WTNT.

During AP, 67/100 of the most highly expressed genes in the MT and the WT were the same in the two strains. Most of the genes (17/33) that were found only on the WTAP top 100 list were ribosomal proteins. Seven others encoded putative SSP effectors, and five had homologs in the PHI database of pathogenicity-associated proteins. Only seven of these 33 genes were statistically more highly expressed in WTAP than MTAP, including five of the SSP effector genes. Many more of the genes found only on the MTAP top 100 list were found in the PHI database (21/33). Two of the genes encoded putative SSP effectors. Only two of the 33 genes were statistically more highly expressed in the MTAP then the WTAP. Neither of these was an SSP.

During BT, 64 genes were shared between the WT and MT top 100 lists. A majority of the genes unique to the top 100 list of the WTBT (22/36) encoded ribosomal proteins. There were also four putative SSP genes, and six that had putative homologs in the PHI database. Only four of the genes were statistically more highly expressed in the WTBT, including one SSP. Among the orphan genes on the MTBT top 100 list were six putative SSP effector genes, and 18 with homologs in the PHI database. Only eight of the genes were statistically more highly expressed in the MTBT, none of which were SSP effectors.

3.4.5.2 Differentially expressed genes. A total of 2412 differentially regulated genes had a log2 fold change ≥ 2.00. There were 760 genes that were differentially expressed during the transition from appressoria to biotrophy (WTAP:WTBT), and 992 during the shift from biotrophy to necrotrophy (WTBT:WTNT)(Tables 2.1, 2.2) . A total of 228 genes were differentially
expressed in both comparisons. Among the genes that were different in only one comparison, 159 “early genes” were significantly higher only in AP, and 440 “late genes” were significantly higher only during NT.

For the MT, only two phases of development occurred in leaf sheaths (AP and BT) (Torres et al. 2013). In contrast with the large change in gene expression in the WT during the transition from AP to BT (760 genes), only 20 genes were differentially expressed between these two phases in the MT, all of them more highly expressed during BT (MTAP:MTBT_down) (Tables 2.1, 2.2). One-third of these genes encoded carbohydrate-active enzymes (CAZymes). Four encoded putative SSP effectors, and ten encoded putative homologs of proteins included in the PHI database.

When comparing the WTAP and MTAP treatments, 218 genes were differentially expressed: 74 were higher in the WTAP (WTAP:MTAP_up); and 144 were higher in the MTAP (WTAP:MTAP_down) (Tables 2.1, 2.2). There were 714 genes that were differentially expressed between the WTBT and MTBT treatments, including 192 that were higher in WTBT (WTBT:MTBT_up) and 522 that were higher in MTBT (WTBT:MTBT_down) (Tables 2.1, 2.2).

Several classes of proteins that could be important in pathogenicity appeared to be over-represented or under-represented, relative to their abundance in the genome, among the differentially-expressed genes. These included genes encoding SSP and SSP-CR, secreted proteases, and carbohydrate-active enzymes (CAZymes) (Table 3.4).

3.4.5.3. Secreted proteins are over-represented among differentially expressed genes. Looking at the secreted proteins RNAseq data, I noticed that there is an over representation of secreted proteins among the differentially expressed genes. Overall in the genome, 14% of the proteins are predicted to be secreted, but among the differentially expressed genes, the secreted proteins represent 30% (Figure 3.13). Among all the SSPs, 250 (~36%) were differentially expressed, and they can be separated into “early” genes versus “late” genes (Table 3.4). This indicates that the effectors are transcribed in “waves”, as seen in other Colletotrichum species (O’Connell et al. 2012; Gan et al. 2013). Of the C. graminicola early genes that are specifically expressed in appressoria, 114
are SSPs. My theory is that the effectors that are expressed early are most likely the ones involved in establishment of biotrophy and host specificity.

3.4.5.4 Comparison of species and genus specific secreted proteins/effectors. Using the genomes from C. higginsianum (O’Connell et al., 2012) and C. sublineola, the 250 differentially expressed C. graminicola SSP genes were classified as shared by all the three species; shared between C. graminicola and C. sublineola; shared between C. graminicola and C. higginsianum; or species specific to C. graminicola. The genes that were species specific to C. graminicola were expressed in the early stages (appressoria and/or biotrophy) 65% of the time, but the genes shared between two or three species were expressed during the early stages less than 50% of the time (Figure 3.14).

3.4.5.5 Expression Patterns of Conserved SSP Effectors. Homologs of previously described effector proteins in other organisms have been identified in C. graminicola. The BAS2 and BAS3 homologs are among the most highly expressed genes throughout development in planta in both the MT and WT, but they are both expressed at significantly lower levels in WTNT when compared to either WTAP or WTBT (Tables 2.1, 2.2)

Two of the CFEM SSP proteins (GLRG_02673 and GLRG_06605) have an early pattern of expression, while GLRG_09687 is expressed later (NT/AP comparison) (Tables 2.1, 2.2).

The two LysM domain SSP proteins, including the homolog of ClH1, are expressed during the early stages of fungal colonization in the WT strain (Tables 2.1, 2.2). Also, the expression of this in MTAP is significantly different than from the WT at the same stage (MTAP/AP comparison) (Tables 2.1, 2.2).

C. graminicola has two chitin-binding domain SSPs (GLRG_06483 and GLRG_10441), and the first one is differentially expressed in the WTBT and WTNT (Tables 2.1, 2.2).

Besides effectors that are thought to be involved in the establishment of biotrophy, hemibiotrophic pathogens also secrete proteins that induce PCD, which are thought to be involved in the switch to necrotrophy. I identified homologs in C. graminicola for all but one of the six NEP proteins identified in C. higginsianum (Table 3.3). In C. higginsianum, ChNLP1 and ChNLP2 are
expressed during the switch from biotrophy to necrotrophy, ChNLP3 and ChNLP5 are expressed in appressoria, and ChNLP4 is not expressed. The ability of ChNLP1 to cause PCD was functionally confirmed, while ChNLP3 did not trigger necrosis. *C. graminicola* homologs of ChNLP2, ChNLP3 and ChNLP5 are differentially expressed. Only CgNLP1 and CgNLP2 in *C. graminicola* share the amino acids residues crucial for NEP activity. CgNLP2 is most highly expressed during WTNT. This gene is also more highly expressed in WTBT vs MTBT. In contrast, the CgNLP3 and CgNLP5 transcripts are more abundant during WTAP, similar to the expression patterns of their homologs in *C. higginsianum*. Neither of these proteins has all the amino acid residues that are essential for induction of PCD (Ottmann et al., 2009). Expression of CgNLP4 was very low *in planta*, similar to its homolog in *C. higginsianum* (Kleeman et al., 2012) (Table 3.3).

There are 17 proteins with the NUDIX domain identified by Pfam in *C. graminicola*. Only one of them (GLRG_05582) was highly expressed in WTBT and WTNT (Tables 2.1, 2.2). None of the NUDIX domain proteins was classified as an SSP.

The homolog of the NIS1 gene in *C. graminicola* (GLRG_05338), was very highly expressed at the same level across all developmental stages of the pathogen, including biotrophy.

3.4.5.6. **Effectors differentially expressed in the mutant.** There were 20 genes that were differentially expressed in the transition from MTAP and MTBT, all more highly expressed in MTBT. Ten of the 20 have homologs in the PHI database. Ten of the 20 are predicted to encode secreted proteins, including 4 SSPs.

There is one glycosyl hydrolase SSP effector that was differentially expressed in the MT, a homolog of XYL2 from *C. carbonum* (GLRG_05524). Work done by Nguyen and collaborators (2011) in *M. oryzae* showed a reduction in the virulence when the homolog of this gene was knocked out in that fungus.

The gene GLRG_06286 encodes a secreted metalloprotease that is homologous to MEP1 from *Coccidioides posadasii* (Tables 2.1, 2.2). These types of protease enzymes require a metal for activity. In *C. posadasii*, MEP1
is secreted by the fungi and prevents host detection by digesting surface antigens from mice cells. When this gene was mutated, virulence was reduced (Hung et al. 2005). There are two additional secreted metalloproteases among the differentially expressed secreted proteins, although they are larger than the 300 amino acid upper cutoff to be called SSP.

One SSP-CR (GLRG_11440) that is found in multiple *Colletotrichum* species, and one orphan SSP (GLRG_03485), are differentially expressed in the MT. GLRG_03485 is also differentially expressed specifically in WTAP. (This gene appears to be missing the 3’end, and is very short – 66 aa/198nt – supercontig1.10, start at 726614).

### 3.4.6 Identification of non-annotated effectors

Kleemann and his collaborators (2012) used data from RNA sequencing to identify 54 putative *C. higginsianum* effector genes that had not been included in the Broad genome annotation. It’s possible that these genes were missed by the Broad annotation because they do not match known genes in the databases, and Broad uses prediction tools that take these prior data into account.

For all my analysis I used the Broad predictions from *C. graminicola* and *C. higginsianum*, but for the *C. sublineola* strain CgSL1 and the *C. graminicola* strain *M5.001* we used the Maker prediction since we did not have access to Broad’s proprietary annotation program Calhoun. The newly published *C. sublineola* TX430BB strain also used the Maker pipeline for genome annotation (Baroncelli et al. 2014). Different gene prediction programs use different algorithms and therefore often yield different annotations.

Broad’s Calhoun protocol uses EST evidence and BLAST homology against Genbank’s NR database, among other criteria, to identify genes. They train the gene prediction program with a combination of other programs such as FGENESH, GENEID, and GeneMark, and EST-based automated and manual gene models. After that, they apply an in-house pipeline to improve their annotation of the genome. A full description can be found at (http://www.broadinstitute.org/annotation genome/colletotrichum_group/GeneFinding.html). This is the prediction used for the published genomes of *C.
graminicola and C. higginsianum. Our lab did not have access to Broad annotations, so we used MAKER, a program from GMOD, to predict C. sublineola putative genes. We also used FGENESH as a third prediction method.

Table 3.5 summarizes and compares the results from each prediction program. The numbers do vary, but when I compared the different C. graminicola predictions using BLASTP, 90.5% of the genes were the same. There were 1,653 new genes predicted by FGENESH that did not match the Broad annotation. Out of those, 411 proteins had less than 30 amino acids and were removed from any further consideration. Some authors use an arbitrary lower cutoff of 100 amino acids to avoid including non-protein coding RNAs in their analyses (Dinger et al., 2008), but recently smaller proteins have been associated with important roles in biological processes (Hanada et al., 2007; Yang et al., 2011), so I chose to only remove proteins smaller than 30 amino acids. This left 1,242 proteins with no matches to the BROAD annotation, including 113 secreted protein genes, 79 of which were SSPs. The majority of those genes were orphan C. graminicola genes, but some had hits in the other two Colletotrichum species (Figure 3.15). Out of those 79, 55 had RNAseq transcript evidence, and some of them were expressed early. The transcriptome data helped me to identify and characterize the expression patterns of these effectors and the genes with transcript evidence will be particularly interesting to include in future research.

3.4.7 Expression of the C. graminicola BAS3 homolog in vitro and in planta.

The expression of the BAS3 homolog of C. graminicola was tested by semiquantitative RT-PCR (Figure 3.16). The housekeeping gene actin was used as the RT-PCR normalization control. I used the same amount of total RNA to make cDNA, but each in planta treatment had different amounts of fungal biomass, so normalizing the data was crucial but, it turned out, rather complicated in practice. There is still some room for optimization in my experimental design. The results showed that BAS3 was not expressed in mycelium in vitro, but that there did appear to be expressed in appressoria produced on Petri dishes. Furthermore, it was expressed in planta during
WTAP and WTBT, but expression was reduced in WTNT (Figure 3.16). The *in planta* results are consistent with our transcriptome data that also show higher expression of this effector early. Many effectors are known to be plant-induced including BAS3 in *M. oryzae* (Mosquera et al., 2009), and so it is not surprising to find that the *C. graminicola* homolog is not expressed *in vitro* in mycelium. It was a little bit surprising to see expression in appressoria produced *in vitro*, which suggests a degree of developmental regulation of this gene in addition to response to plant signals. This method has potential for evaluating relative expression patterns of other candidate effectors in future.

3.4.8 Identification of candidate effectors for future studies

Based on my assumption that effectors most likely to be involved in biotrophy would be expressed early, and would be highly divergent, I chose a group of effectors that met these criteria as promising candidates for further study (Figure 3.17). There are 11 orphan *C. graminicola* effectors that are specifically expressed early during infection (Figure 3.17). One of those genes was also identified in the LCD data as being highly expressed *in planta* versus *in vitro*. Some of these effectors are very highly expressed. It would be good to confirm their expression patterns using semi-quantitative PCR and *in planta* reporter gene and localization studies, as well as knockout analyses.
3.5 Discussion

In the work described in this chapter, I have used a bioinformatic (“in silico”) approach to characterize the putative effectorome of *C. graminicola*. In the first part I used a comparative approach to identify putative effectors that are divergent, and thus most likely to be under strong selection pressure, and in the second part of the work I made use of the transcriptome data to identify putative effectors that are expressed early, during the establishment of biotrophy. Finally, I combined these two lists to identify the most promising early, divergent candidate effectors, so that in the future the hypotheses that these are involved in the establishment of biotrophy in *C. graminicola*, and that they are deficient in our CPR1 MT, can be addressed. I had postulated that there would be significant overlap between these two lists if my assumptions were valid, and in fact I did see a trend in which more divergent effectors tended to be expressed earlier while more conserved ones were expressed later. This was not, however, an absolute rule and there were many exceptions, suggesting that my assumptions are probably over-simplified.

No less important, I believe, than generating this list of candidate effectors, was that I also organized and summarized a very large amount of genetic and transcriptomic data in a format that will hopefully make it much more convenient for future researchers (especially those who may not be experienced with programming) to access and use.

*C. graminicola* was one of the last fungal genomes to be sequenced by using primarily Sanger technology. When compared to next generation sequencing (NGS) methods, Sanger sequencing offers the possibility for a higher quality assembly because of the longer sequences that can be generated (Liu et al., 2012). However, Sanger sequencing is relatively expensive and as assembly programs for NGS have improved, microbial genome sequencing projects have largely moved away from Sanger and even 454 in favor of Illumina and other, even cheaper alternatives. The high quality of the *C. graminicola* genome makes it valuable as a reference sequence for other *Colletotrichum* genomes that are sequenced using NGS (e.g. Rech et al., 2014).
The genomes I compared in my work were prepared using different methods for sequencing, as well as for assembly and annotation. Inevitably this will have resulted in some errors in my comparisons, in which genes that appear to be highly divergent may simply have been mis-annotated, mis-assembled, or missing from the sequencing data. I found some apparent examples of this in the secretory pathway of \textit{C. gloeosporioides} (Chapter 2), but it’s harder to recognize this kind of error for effector genes, which are much less conserved in general than other types of genes. I saw that using different annotation protocols produces different results, sometimes VERY different, for the same genome assembly. For example, by using a different gene prediction program, I identified 79 additional putative effector proteins that had not been annotated by the Broad prediction program. I was able to validate most of these novel effectors by using the transcriptome data.

One major challenge in comparative genomics studies is to identify homologous proteins. It is important that proteins being analyzed are truly orthologs rather than paralogs, if the assumptions of the analysis are to be met. I used two different approaches. OrthoMCL is more commonly used, but it failed to include proteins that were more divergent, and the effectors are over-represented in that group (Figure 3.4). Therefore, I also used RBH, which gave me more homology results than OrthoMCL, but has the potential for assigning orthology status to genes that are actually distant paralogs instead. I was encouraged that the two methods agreed more than 90% of the time. When they did not agree, it is likely that OrthoMCL, with its additional weighting steps gives more accurate results. However, for the large number of proteins, many of them putative effectors that OrthoMCL did not classify, RBH was the best method I could use.

Errors like these seem to be an inevitable problem of comparative bioinformatics studies. Unfortunately, there is very little validation or comparison among methods in the literature that can help us to evaluate which is likely to perform the best, and even the best still have weaknesses. It is important to emphasize that all results of my work here should be treated as hypothetical models to be tested by experimentation.
Having said that, comparative bioinformatics, particularly of closely related species, is still a powerful tool to develop new hypotheses regarding species-specific characteristics. In closely related species that differ in pathogenicity, regions of the genome that are unique or highly divergent could be important for those differences (Stukenbrock 2013). My comparative synteny analysis revealed that the genomes of *C. graminicola* and *C. sublineola* are largely co-linear, reflecting their very close evolutionary relationship. I observed some chromosome inversions that seem to have occurred between the two species (Figure 3.2B). Examples include regions of *C. graminicola* chromosome 1 and *C. sublineola* scaffold 5, or of chromosome 3 and scaffold 8. This phenomenon is called mesosynteny (Hane et al. 2011). It has been hypothesized that mesosynteny can be important in speciation (Stukenbrock 2013). I also observed small regions that lacked synteny embedded in the larger co-linear segments. Some of these regions contained putative effector genes. The Symap platform that I used for these comparisons has been set up to allow detailed analysis and visualization of these microsyntenies for future researchers.

The overall degree of similarity of proteins between *C. graminicola* and *C. sublineola* was very high, compared to other species. However, the secretome was less conserved, and SSPs in particular were even less conserved (Figure 3.9, 3.10). Schirawski et al. (2010), comparing closely related corn smut fungi, also found that secreted proteins are more divergent than total proteins. Genes under a high rate of evolution show less amino acid similarity, so I hypothesize that secreted proteins, and especially SSPs, are changing faster. Some of the orphan genes among the SSPs could be involved in the host specificity we observe between *C. graminicola* and *C. sublineola*, and also in early events in pathogenicity in their hosts including establishment of biotrophy. Given this, I was quite surprised to find that relatively few of the proteins (<10%) that were not conserved between *C. graminicola* and *C. sublineola* were predicted to be secreted. Instead, most seemed to be targeted either to the nucleus or to the mitochondria. I’m not sure about the significance of this, and it could simply be due to incorrect calls by the Wolf PSORT protein localization prediction program. It would be good in future to test this experimentally.
The effectorome of *C. graminicola* is not terribly large, by the standard of most plant pathogenic fungi, and it is smaller than that of its close relative *C. sublineola*. There appears to be relatively little evidence for the presence of large effector gene families in *C. graminicola*. One rare example, with five members in *C. graminicola*, is depicted in Figure 3.11A. I identified five genes from *C. sublineola* that appear to belong to the same family. The family is characterized by the presence of six conserved cysteine-rich regions. Interestingly, there is also a family with members limited to *C. graminicola*, *C. sublineola* and *M. oryzae* (Figure 3.11C). These three fungi share a very similar hemibiotrophic lifestyle on graminaceous hosts, so it is possible that this gene family may play a unique role in that lifestyle.

Comparison of two strains of *C. graminicola*, one from North America and one from South America, collected more than a decade apart, revealed very little genome diversity, including among the effectors: the two strains differed in only five SSP proteins. This might suggest that there is little selective pressure driving effector diversification in *C. graminicola*, perhaps because its host, maize, is a cultivated crop with a relatively low level of genetic diversity, in which resistance to *C. graminicola* is primarily due to quantitative trait loci rather than major “R” genes. A recent paper by Rech et al. (2014) also reported very low levels of diversity among effector proteins across seven different strains of *C. graminicola* from a worldwide collection. However, they found evidence for diversifying selection in the 5’UTR regions of the effector genes, and they suggested that differences in effector expression may be more important than differences in protein sequence in considering effector evolution and selection. I did not investigate 5’UTR sequences in my study, but it would be a good thing to do in future.

I found no evidence for a widely-conserved sequence motif like RXLR (Birch et al. 2008; Morgan and Kamoun 2007) among my effectors (Figure 3.12). There were a few motifs, including several that were cysteine-rich, that were shared among smaller groups of proteins. Cysteine forms disulfide bonds, thus potentially stabilizing protein tertiary structures important for function, or protecting them from host proteases. Effectors with the same motifs might have similar functions. The transcriptome data shows that some groups of proteins
that share the same motif are similarly regulated. For example, among the 10 sequences with Motif07, eight are differentially regulated, and seven are expressed preferentially early during infection. However, most of the motifs don’t seem to match particular patterns of expression. The significance of these motifs, if any, would need to be tested experimentally in mutagenesis studies.

I used all the data from my genome comparison of *C. graminicola* and its close relative *C. sublimeola*, together with the transcriptome data from pathogenic and non-pathogenic strains of *C. graminicola*, to identify a group of effector candidates that I hypothesize are most likely to be involved in the species-specific establishment of biotrophy and suppression of PCD. Dr. M. Torres, in her dissertation research, reported that the development of the MT is stopped very early during biotrophy in the host tissues, and thus genes that are differentially expressed in MTBT are likely to be those that are turned on first during the transition from appressoria to biotrophy (Torres, 2013). These genes are also good candidates for biotrophy determinants.
Figure 3.1 Bioinformatic pipeline for prediction of the *Colletotrichum* effectorome. The pipeline is composed of the major steps used in the characterization of the proteins.
Figure 3.2 Synteny between *C. graminicola* and *C. sublineola*. A) Global view of syntenic alignments between *C. graminicola* chromosomes (green), and scaffolds of *C. sublineola* (purple) and *C. higginsianum* (grey). B) Micro synteny between each *C. graminicola* chromosome (vertical axis) and *C. sublineola* supercontigs (horizontal axis). Homologous regions are identified inside the boxes.
Figure 3.3 Amino acid similarity of orthologous proteins between *C. graminicola* and other sequenced *Colletotrichum* species.
Figure 3.4 Orthology by OrthoMCL. Phylogenetic tree (unscaled) based on NCBI Taxonomy Browser. OrthoMCL data provided by Dr. Neil Moore showing the number of shared OrthoMCL groups among all species, as well the number of non-conserved genes that had no homologs between the ten species.

<table>
<thead>
<tr>
<th>Name</th>
<th># OrthoMCL groups</th>
<th>Non-conserved Genes</th>
<th>Total Genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspergillus flavus</td>
<td>7870</td>
<td>4364</td>
<td>13487</td>
</tr>
<tr>
<td>Magnaporthe oryzae</td>
<td>8102</td>
<td>6728</td>
<td>15485</td>
</tr>
<tr>
<td>Epichloë festucae</td>
<td>6100</td>
<td>2001</td>
<td>8356</td>
</tr>
<tr>
<td>Trichoderma reesei</td>
<td>7651</td>
<td>1077</td>
<td>9129</td>
</tr>
<tr>
<td>Fusarium graminearum</td>
<td>10258</td>
<td>3077</td>
<td>14038</td>
</tr>
<tr>
<td>Fusarium oxysporum</td>
<td>11224</td>
<td>6543</td>
<td>20925</td>
</tr>
<tr>
<td>Verticillium dahliae</td>
<td>8366</td>
<td>1643</td>
<td>10535</td>
</tr>
<tr>
<td>Colletotrichium higginsianum</td>
<td>10140</td>
<td>4711</td>
<td>16150</td>
</tr>
<tr>
<td>Colletotrichium graminicola</td>
<td>10405</td>
<td>1106</td>
<td>12006</td>
</tr>
<tr>
<td>Colletotrichium sublineolium</td>
<td>10472</td>
<td>2200</td>
<td>13331</td>
</tr>
</tbody>
</table>
Figure 3.5 Orthology by Reciprocal Best Hit. Each diagram shows the orthology between the main species (on the top) with the other two.
Figure 3.6 Carbohydrate-degrading enzymes (CAZymes) in *C. graminicola* (Cgram), *C. sublineola* (Csub) and *C. higginsianum* (Chig), separated by different classes.
Figure 3.7 Blast results of proteins shared between *C. graminicola* and the other two species using NCBI database.
Figure 3.8 Orthology among the small secreted proteins between all 3 *Colletotrichum* species. Each diagram shows the orthology between the main species (on the top) with the other two.
Figure 3.9 Amino acid similarity of orthologous proteins between *C. graminicola* and the other two species, *C. sublineola* (A) and *C. higginsianum* (B). Proteins were separated into not secreted versus secreted proteins (SP) categories.
Figure 3.10 Amino acid similarity of orthologous proteins between *C. graminicola* and the other two species, *C. sublineola* (A) and *C. higginsianum* (B). Only secreted proteins were considered, separated between those that had more than 300 aa (SP) and the small secreted proteins (SSP).
Figure 3.11 Alignment of effector protein families.
Figure 3.12 Sequence of motifs discovered on SSPs of *C. graminicola* using MEME. Number of proteins with the motif sites identified are shown in each figure.
Figure 3.13 Comparison between secreted proteins present in the genome and secreted proteins present in differentially expressed RNAseq transcripts, showing the over-representation of secreted proteins *in planta.*
Figure 3.14 Timing of the differentially expressed SSPs based on their homology to other *Colletotrichum* species.
Figure 3.15 Identification of non-annotated effectors in *C. graminicola* using FGENESH prediction programs.
Figure 3.16 Expression of *C. graminicola* BAS3 on *in planta* and *in vitro* conditions. AP: appressoria, BT: biotrophic, NT: Necrotrophic, IVAP: *in vitro* appressoria, Fries: minimal Fries medium, PQ: Paraquat, TUN: tunycamycin.
Figure 3.17 Orphan candidate effectors highly expressed in the early stages of infection.
Chapter 4

Genomic and functional analyses of stress response in wild type and mutant strains of *Colletotrichum graminicola in vitro* and *in planta*

4.1 Overview

There is evidence in the literature that plant tissues that are actively defending themselves produce a stressful environment for the pathogen (Torres 2010). In this chapter, I investigated the nature of stress responses in WT *C. graminicola in planta*, and under chemically imposed stress *in vitro*. I expected to receive some insights from this work into the types of stress that are being experienced by the WT during various phases of development *in planta*. The MT is able to grow normally in maize tissues that are not alive and actively defensive. In this chapter I addressed the hypothesis that *C. graminicola* is exposed to stresses *in planta*, and that the MT is deficient in its ability to adapt to those stresses by testing three predictions related to this hypothesis. My first prediction was that the MT would be more sensitive than the WT to stress *in vitro*. My second prediction was that WT and MT strains would express stress response genes *in planta*. My third prediction was that the MT would differ from the WT in the expression of stress response genes *in planta* and *in vitro* under stress conditions.

4.2 Introduction

4.2.1. Fungal Response to Stress

All organisms, including fungi, have the capacity to adapt to environmental stresses via the activation of a variety of stress response signaling pathways (Kültz 2003). The functions of these pathways and their component stress response proteins have been characterized in detail in several model organisms, including bacteria, mammals, plants, and fungi. Recent comparisons of these models with related species that have been characterized by genomic analyses have shown a generally high level of conservation of the pathways, although they often differ in some details (Smith et al. 2010). The
general strategy of the stress response pathways is activation of a variety of sensors in response to stress, which in turn initiate secondary messenger signaling pathways, typically MAP kinase cascades, which serve to amplify the signals, and finally triggering transcription factors that regulate multiple genes involved in protection and adaptation to the individual stressors.

Results of a comparative genomics study of a large number of species across several different kingdoms showed that 67 of the 368 phylogenetically most highly conserved proteins (or almost 1 in 5) were involved in stress response (Kültz 2003; Kültz 2005). In addition to conservation of the coding regions, stress response genes typically also have conserved upstream regulatory elements that bind to the transcription factors that are controlled by stress response pathways. Stress responsive upstream elements are conserved in a wide range of species, from yeast (Estruch 2000; Hohmann 2002) to humans (Papadakis and Workman, 2014) and plants (Naika et al., 2013). Some conserved sequences located in the 3'UTR have been shown to be involved in post-transcriptional regulation of gene expression in response to stress (Spicher et al. 1998). Alternative splicing of transcripts is known to occur in response to stress. In yeast, alternative splicing was especially common in the transcripts of ribosomal proteins in response to nutritional stress (Bergkessel et al., 2011). Stress response can also be regulated post-translationally by control of protein synthesis rates, release of membrane-bound regulatory molecules, and changes in locations of proteins as compartments are disrupted by stress (Kaufman et al. 2002; Kaufman 1999; Guerra et al. 2015; Bernales et al. 2006; Kültz 2005).

Different stress conditions frequently elicit a response from the same genes, demonstrating that stress pathways are highly interconnected (Chasman et al. 2014; Bergkessel et al. 2011; Breitkreutz et al. 2010). The crosstalk among different stress pathways allows a stress signal to be amplified, and activation of multiple pathways in response to stresses can increase pathogen survival (Fuchs and Mylonakis, 2009; Hayes et al., 2014).
4.2.2 The Role of Stress Response in Pathogenicity

Evidence suggests that the ability to cope with external stresses is crucial for a pathogen to successfully colonize and complete its life cycle in a host (Chung, 2012; Doehlemann and Hemetsberger, 2013). The host environment is believed to be very stressful, due to pre-formed and inducible defensive mechanisms deployed by the host to protect itself from pathogen attacks. Several stress response pathways have been specifically implicated in pathogenicity.

4.2.2.1 Secretion Stress: Transported proteins are processed by the signal peptidase complex (SPC) and enter the lumen of the endoplasmic reticulum (ER), where they are folded into their proper conformations through the activities of various chaperone proteins and enzymes. They are then transported in vesicles from the ER to the Golgi, where they are often further modified, e.g. by adding glycosyl groups, before being directed to their final internal or external locations (reviewed by Vitale and Denecke 1999: also see Chapter 2 of this dissertation). Secretion stress occurs when this process of protein transport is perturbed, resulting in a potentially lethal accumulation of misfolded proteins in the lumen of the ER. The Unfolded Protein Response (UPR) is a conserved stress response pathway that helps the cell to maintain essential transport functions and viability in the presence of secretion stress. Activation of UPR results in removal of misfolded proteins, decreases in the production and transport of non-essential proteins, and adaptive increases in overall secretory capacity (Heimel, 2015; Hollien, 2013; Lai et al., 2007).

The UPR pathway has been most closely studied in the budding yeast *Saccharomyces cerevisiae*, but it seems to be highly conserved in all eukaryotes (Arvas et al., 2006; Lin et al., 2008; Richie et al., 2009; Travers et al., 2000). The signal to activate the UPR originates in the ER and travels to the nucleus, resulting in an increase in the expression of chaperones and folding proteins (Patil and Walter, 2001). Ire1 is an ER transmembrane protein that interacts with BiP, (aka Kar2 in yeast), a member of the Hsp70 family, on the luminal side of the membrane. BiP is the most abundant chaperone protein in the ER and is highly conserved across a wide range of organisms (Pincus et al. 2010; Gething 1999). BiP is involved in stabilization of nascent proteins as they
pass through the translocon into the ER lumen, and it binds to proteins to facilitate their correct folding. Currently there are two proposed models of how UPR is activated (Guerriero and Brodsky 2012). In the first one, when misfolded proteins accumulate, Ire1 transduces a UPR signal across the membrane, causing BiP to dissociate from it and to bind to unfolded proteins. This also activates the endoribonuclease function of Ire1 which catalyzes the splicing of the transcription factor Hac1 pre-mRNA. Splicing of the Hac1 transcript activates its regulatory activity and allows Hac1 to switch on genes that have a conserved UPR element in their promoters (Mori et al., 1996; Mori et al., 1998). These genes are involved in moderating secretion stress, and include various heat shock proteins (Miskei et al., 2009; Schröder and Kaufman, 2005; Gülow et al., 2002). The other model was proposed by Pincus and co-authors (2010). In this model, BiP binds to and inactivates Ire1 during low stress conditions. Binding of unfolded proteins to Ire1 causes the formation of Ire1 complexes, releasing it from BiP, and resulting in its activation to process the Hac1 pre-mRNA (reviewed by Gardner et al. 2013).

The UPR is a central stress response pathway that is integrated closely with responses to a large number of other environmental stresses, many of which directly or indirectly cause perturbation of protein transport and an accumulation of misfolded proteins. In plants it appears that proteins that detect secretion stress and trigger UPR are adapted as general stress surveillance molecules that also activate other stress response pathways (Liu and Howell 2010; Takato et al. 2013). UPR can be induced in vitro by treatment with chemicals that interfere with protein folding or modification (Denecke et al., 2012; Guillemette et al., 2007; Iwata et al., 2010; Satpute-Krishnan et al., 2014; Yi et al., 2009). Two commonly used chemicals are Dithiothreitol (DTT) and tunicamycin. DTT reduces disulfide bond formation, while tunicamycin blocks the synthesis of N-linked glycoproteins.

Numerous human diseases are linked to pathologic conditions that trigger secretion stress, including diabetes, some types of cancer and viral infections, and neurodegenerative diseases such as Alzheimer’s (Hutt et al., 2009; Kadowaki and Nishitoh, 2013; Lin et al., 2008; Myung et al., 2001; Schröder and Kaufman, 2005).
The UPR and the secretory system have been directly implicated in fitness and virulence of filamentous fungi. For example, virulence of the opportunistic human pathogen *Aspergillus fumigatus* has been associated with activation of the UPR and the endoplasmic reticulum-associated degradation pathway (ERAD), both related to secretion stress (Richie et al. 2011; Feng et al. 2011). Disruption of Hac1, the major transcriptional regulator of UPR, produced an *A. fumigatus* mutant that was reduced in virulence in mouse models, with increased sensitivity to heat stress and fungicides, and an inability to assimilate nutrients from complex substrates (Richie et al. 2009). Growth of the Hac1 mutant was comparable to the wild-type strain under normal conditions, but conidiation was reduced (Richie et al. 2009). In the necrotrophic Brassica pathogen *Alternaria brassicicola* the disruption of a homolog of Hac1 resulted in a nonpathogenic mutant which also had a cell wall defect and a reduced capacity for secretion (Joubert et al. 2011). In a study in *Magnaporthe oryzae*, the *Lhs1* chaperone which is important for protein transport into the ER and proper folding, was essential for pathogenicity (Yi et al. 2009).

4.2.2.2 Oxidative Stress: The production of reactive oxygen species (ROS), the so-called “oxidative burst”, is one of the earliest observable plant defense responses to pathogen attack (Apel and Hirt 2004; Dickman and de Figueiredo, 2013; Moye-Rowley, 2003). ROS are generated during normal cellular metabolism, but they are transient and tightly regulated in order to maintain them at tolerable levels (Angelova et al., 2005). Exposure to high levels of ROS during the oxidative burst damages proteins, lipids and nucleic acids of both plant and pathogen, and can result in programmed cell death (PCD) of the plant cells (Ikner and Shiozaki, 2005). PCD is beneficial to necrotrophic pathogens that can only colonize dead host cells (Chung, 2012; Dickman and de Figueiredo, 2013). Fungi react to ROS exposure by activating the oxidative stress response pathway, leading to the production of protective antioxidant molecules such as glutathione and thioredoxin, and enzymes such as catalases, peroxidases, and superoxide dismutases that specifically degrade and detoxify ROS (Montibus et al., 2013). Oxidative stress can be induced in vitro by treatment with various chemicals including menadione, hydrogen peroxide (H₂O₂), or the herbicide Paraquat (Kavitha and Chandra 2014;
Angelova et al. 2005). Menadione induces oxidative stress by forming superoxide anions, and Paraquat generates reactive oxygen species (ROS) by the induction of electron transfer in multiple subcellular compartments (Abegg et al., 2011; Angelova et al., 2005; Fernandes et al., 2007). Different chemical treatments sometimes induce different responses: thus H$_2$O$_2$ induced production of both superoxide dismutases and catalases in several different fungi, but Paraquat only induced superoxide dismutases (Angelova et al., 2005). However, in a different study, genes induced by menadione and H$_2$O$_2$ showed significant overlap (Jamieson 1992). Moreover, in *M. oryzae*, transcription factors upregulated during *in vitro* oxidative stress were similar to ones induced *in planta* (Park et al. 2013).

Several fungal genes related to oxidative stress and ROS detoxification were induced during infection of chickpea by the necrotrophic fungus *Ascochyta rabiei* (Singh et al., 2012). A glutathione peroxidase mutant of *M. oryzae* produced lesions in barley that were smaller than those produced by the WT (Huang et al., 2011a, 2011b). Mutation of an ortholog of the yeast oxidative stress transcription factor Yap1 in *Alternaria alternata*, a necrotroph that causes citrus brown spot, resulted in a loss of pathogenicity and reduced expression of catalases, peroxidases, and superoxide dismutases (Lin et al., 2009). Oxidative stress has been linked to the production of toxic secondary metabolites in fungi including *Fusarium graminearum* (Ponts et al., 2006) and *M. oryzae* (Forlani et al. 2011).

4.2.2.3 Osmotic Stress: Osmotic stress occurs due to an imbalance in the solute concentration between the internal and external cell environment. If the cell occupies a hyperosmotic environment, it will lose water resulting in plasmolysis. In hypo-osmotic conditions the cell will absorb water, and can eventually burst. Osmotic stress can be induced *in vitro* by exposure of the fungi to high levels of an osmolyte such as sorbitol or sodium chloride (Kovács et al., 2013; Rispail and Pietro, 2010). Osmotic stress causes impairment of cell function, ultimately resulting in cell death. Cells protect themselves by activating the osmotic stress pathway, which allows them to regulate concentrations of internal solutes such as glycerol to maintain osmotic balance (Posas et al. 1996). Fungi can adapt even to sudden changes in external osmolarity. In
*Candida albicans*, which causes Candidiasis in mammals, this pathway has been shown to be very important for pathogenicity. The *Hog1* kinase is a highly conserved protein that when mutated causes inability to grow under osmotic stress conditions in *S. cerevisiae* (Brewster et al., 1993). Deletion of the *Hog1* homolog in *C. albicans* results in a reduction in virulence (Cheetham et al. 2011). Interestingly, the *C. albicans* *Hog1* mutant does not become more sensitive to osmotic stress, suggesting that it has an additional osmotic stress response pathway that yeast lacks (Enjalbert et al. 2003).

In some filamentous fungi, activation of *Hog1* results in accumulation of molecules, including saccharides and glycerol, which diminish osmotic stress. However, in these fungi *Hog1* is also an important regulator of the oxidative response pathway (Bilsland et al., 2004). When the *Hog1* ortholog of *Aspergillus nidulans* was deleted, it resulted in reduced conidiospore viability and increased sensitivity to oxidative stress and heat shock (Kawasaki et al., 2002). Similar defects were found in *Hog1* mutants in *F. graminearum* (Zheng et al. 2012), which also became less pathogenic to wheat. In *Alternaria alternata*, deletions of *Hog1* resulted in increased susceptibility to oxidative and salt stress and loss of pathogenicity on tangelo leaves (Chung 2012). *Ssk1* is a critical activator of *Hog1* in yeast but when *Ssk1* was mutated in *Aspergillus* there were no changes in sensitivity to osmotic stress (Duran et al. 2010). This further illustrates differences that exist between yeast and other fungi in the osmotic stress response pathway (Kültz and Burg 1998; Miskei et al. 2009).

4.2.2.4 Cell Wall Stress: Fungi have wall sensors that detect external stresses and respond to them via the cell wall integrity pathway, which triggers adaptive changes in the composition and strength of the wall (reviewed by Fuchs and Mylonakis 2009). The fungal cell wall serves as the primary sensory connection between the fungal organism and the external environment, and the cell wall integrity pathway thus serves as a central regulatory pathway that interfaces with other response pathways to several different external stresses including osmotic, pH, thermal, oxidative, and nutrient stresses. This ensures that the critical functions of the cell wall can be maintained and optimized under a range of stress conditions (Fuchs and Mylonakis 2009; Nikolaou et al. 2009). The fluorescent dye Calcofluor can be used to induce cell wall stress in fungi.
Calcofluor interacts with chitin, and interferes with cell wall assembly (Kovács et al. 2013).

In the cell wall integrity pathway, cell wall stress is first detected by membrane receptors. Mutation of the \textit{Wsc1} cell wall stress sensor gene in yeast results in cell rupture at elevated temperatures (Gray 1997). The wall sensors recruit Rom, which in turn triggers the small G protein Rho, which initiates activation of a MAP kinase cascade. Mutations in either Rom or Rho cause wall and growth defects in yeast (Schmelze 2002). Mutations of the kinases also result in sensitivity to cell wall stress and growth defects in yeast and in \textit{C. albicans} (reviewed by Levin 2005, Blankenship 2010; Navarro-Garcia 1995). Homologs of these kinases in filamentous fungi have been shown to be involved in cell wall structure, conidiation, and pathogenicity (Zhao et al., 2007). The cell wall integrity pathway terminates with activation of several transcriptional regulators that have multiple targets, including genes involved in chitin synthesis. The GFA (fructose-6-phosphate amidotransferase) enzyme, which is responsible for the first, rate-limiting step in chitin synthesis, is highly induced under cell wall stress conditions in yeast (Lagorce et al. 2002) as well as in filamentous fungi like \textit{A. niger} and \textit{F. oxysporum} (Ram et al. 2004). The result of activation of the cell wall integrity pathway is an increase in chitin accumulation, and stiffening and strengthening of the cell wall (Dallies 1998).

There is evidence that the cell wall integrity pathway plays a role in pathogenicity. Disruption of the homolog of the \textit{Slit2} MAPK in the entomopathogenic fungus \textit{Beauveria bassania} resulted in alterations in the cell wall, hypersensitivity to the cell wall inhibitory compound Congo Red, and a reduction in conidiation and virulence (Luo et al., 2012). Disruption of the \textit{Bck1} MAPKKK homolog of the mycoparasite \textit{Coniothyrium minitans} also resulted in wall and conidiation defects, and reduced virulence of this parasite to its host \textit{Sclerotinia sclerotiorum} (Zeng et al., 2012). A \textit{F. oxysporum} mutant with a defect in the Rho1 GTPase also showed cell wall defects, and was less virulent to tomato plants (Martínez-Rocha et al. 2008).
4.2.3 Evidence that Cpr1 and the signal peptidase may play a role in stress response

To successfully infect a plant, a pathogenic fungus must secrete an array of proteins that promote susceptibility and facilitate nutrient uptake from the host (Tang et al. 2006). The ER plays a central role in protein secretion, and the SPC is the gateway to the ER for secreted proteins. A sudden increase in secretory activity can be triggered during differentiation of specific secretory cells, or in response to various stresses that may be encountered in planta (Kaufman et al., 2002). An increase in secretory activity during the establishment of biotrophy and the switch to necrotrophy could lead to secretion stress in *C. graminicola*, which would be expected to trigger the UPR (Richie et al., 2009; Schröder and Kaufman, 2005).

*Cpr1* has been specifically connected to secretion stress responses in fungi. In *A. niger*, a 7-fold increase in the rate of translation of the *Cpr1* homolog was reported in response to secretion stress induced by chemicals (Guillemette et al., 2007). Another study reported that several genes encoding proteins in the secretory pathway, including *Cpr1*, were up-regulated in *A. niger* under conditions of carbon starvation stress *in vitro* (Jørgensen et al., 2009). Other members of the signal peptidase complex (SPC) have also been implicated in stress response. The Spc2 protein of the yeast SPC is not essential, but it is required for full enzymatic activity of the SPC *in vitro* (Wolfram Antonin, Meyer, and Hartmann 2000). Mullins and colleagues found that an Spc2 mutant accumulated unfolded and unprocessed proteins in the ER under temperature stress (Mullins et al. 1996), suggesting that Spc2 is required for optimization of protein transport during stress.

In this chapter, I explore a possible connection between *Cpr1* and stress response in *C. graminicola*. My hypothesis is that *C. graminicola* is exposed to stresses in planta, and that the MT is deficient in its ability to adapt to those stress. To address this hypothesis I tested three predictions: A) the MT will be more sensitive than the WT to stress *in vitro*; B) WT and MT strains express stress response genes *in planta*; and C) the MT will differ from the WT in the expression of stress response genes *in planta* and *in vitro* under stress conditions.
4.3 Material and Methods

4.3.1 Strains and culture conditions

The *C. graminicola* strain M1.001 isolated from diseased maize was obtained from the late Dr. Robert Hanau (Purdue University, West Lafayette, IN, U.S.A.). The nonpathogenic mutant strain (MT) and its complement (MT-C), described in Thon et al. (2002), were both derived from M1.001 (WT). All isolates were routinely cultured on Potato Dextrose Agar (Difco Laboratories, Detroit, PDA) at 23°C under continuous fluorescent light. Spores were collected and used for inoculations as described in Chapter 2 of this dissertation.

4.3.2 Pharmacological analysis of WT and MT stress responses *in vitro*

The effect of various stresses on linear growth of the MT, WT, and MT-C strains was determined by using race tube assays as described in detail in Appendix III of this dissertation. The tubes contained minimal Fries medium (30 g sucrose, 5 g ammonium tartrate, 1.0 g ammonium nitrate, 1.0 g potassium phosphate, 0.48 g magnesium sulfate anhydrous, 1.0 g sodium chloride, 0.13 g calcium chloride/liter of H$_2$O) solidified with 1.5% agar. After autoclaving and cooling, the medium was amended with stress-inducing chemicals as described below. Controls in each case contained the same minimal Fries medium amended with the same concentrations of the dilution buffer only. Linear growth was measured after 10 days of incubation at 23°C.

DTT and tunicamycin were used to induce secretion stress. DTT was diluted in water and added to the media to produce final assay concentrations of 2.5, 5, 10, 15 and 20 mM. A stock of 1 mg/ml tunicamycin was prepared in DMSO, and the stock was diluted and added to the media to produce the final assay concentrations of 2, 4, 6, 8 and 10 µg/ml. Paraquat and menadione were used to induce oxidative stress. Paraquat assay concentrations were 0.5, 1.0, 3.0, 5.0 and 10 mM, and menadione assay concentrations were 50, 100, 200, 300 and 500 µM. Sorbitol and sodium chloride (NaCl) were used to induce osmotic stress. Sorbitol was assayed at concentrations of 0.4, 0.8, 1.2, 1.6 and 2 M, and NaCl concentrations were 0.2, 0.4, 0.8, 1.2, and 1.6 M. To induce cell wall stress, I used Calcofluor at 1, 1.5, 2, 2.5 and 3 mg/ml.
To evaluate the effect of temperature stress, growth on unamended minimal Fries medium at 18, 30 and 37°C was compared with the control that was cultured at the optimum temperature of 23°C.

To evaluate the effect of nutritional stress, the strains were grown in a carbon and nitrogen-limited minimal Fries medium. The control was minimal Fries medium as described above, and the treatments consisted of Fries minimal medium containing one half, one quarter, and one eighth the normal concentration of carbon (sucrose) and nitrogen (ammonium nitrate and ammonium tartrate) present in the original formula. For pH experiments, minimal Fries medium was adjusted to pH values of 4, 5.5, 8, and 10. The control was the unadjusted pH 6.0 medium.

The effect of each treatment on fungal growth was assessed by calculating the percentage of growth of each strain relative to their non-stress control for each chemical. Each treatment had four repetitions and most experiments were repeated at least three times. Experiments were not repeated if the first repetition clearly showed that there was no difference in the WT vs MT response. To compare the stress sensitivity of each fungal strain the mean relative growth (%) was calculated for each species under the conditions tested. To measure relative growth, the amount of growth in presence of stress was divided by the amount of growth observed for unstressed cells of the same species, and expressed as a percentage.

4.3.3 In silico identification of C. graminicola stress response genes and pathways

I compiled a list of candidate stress response genes in C. graminicola by using BLASTP, with an e-value of 1e-5, to identify putative homologs of stress-associated sequences that had been described in the literature from S. cerevisiae, A. niger, Trichoderma reesei and M. oryzae (Guillemette et al. 2007; Arvas et al. 2006; Mathioni et al. 2011; Jørgensen et al. 2009; Nikolaou et al. 2009). Additional stress response genes were identified by searching for homologs of genes included in the Fungal Stress Response Database (FSRD) (www.http://internal.med.unideb.hu/fsrd/). The FSRD contains 1985 genes with
verified functions in stress response in fungi, including pathogens of humans and plants, as well as species of industrial significance (Karányi et al., 2013).

To identify UPRE (UPR elements), I searched for motifs that had previously been described in other organisms (Kokame et al. 2001; Fordyce et al. 2012; Gilchrist et al. 2006; Roy and Lee 1999) in the upstream region of selected genes involved in stress response (Table 4.1). The motifs included sequences that were identified in both mammalian systems and yeast: 5'-CCAATNN5CCACG-3'; 5'-ATTGGNN5CCACG-3'; 5'-GGCCAGCTG-3'; 5'-CAGcGTG-3'; 5'-TACGTG-3'; 5'-AGGACAAC-3'.

Putative stress response pathways for C. graminicola were constructed based on pathways that had been described for model fungi in the literature (Chen and Dickman, 2004; Geysens et al., 2009; Ikner and Shiozaki, 2005; Jørgensen et al., 2009; Mathioni et al., 2011; Miskei et al., 2009; Moye-Rowley, 2003; Nikolaou et al., 2009). Putative C. graminicola homologs of stress response genes included in each pathway were identified by BLASTP as described above. To identify the degree of conservation of the proteins associated with these pathways, five species of Colletotrichum were compared with S. cerevisiae, as described in Chapter 2.

4.3.4 Transcriptome analysis

A description of the experimental and statistical analysis protocol for the in planta transcriptome study was presented in Chapter 2 of this dissertation.

I identified genes predicted to be associated with stress from among the 100 most highly expressed transcripts for each treatment. As detailed in Chapter 2, the in planta treatments were: WT pre-penetration appressoria (WTAP); WT biotrophic phase (WTBT); WT necrotrophic phase (WTNT); MT pre-penetration appressoria (MTAP); and MT biotrophic phase (MTBT). I also identified stress genes from among the list of genes that were differentially expressed in each comparison. As detailed in Chapter 2, the comparisons were WTAP:WTBT; WTAP:WTNT; WTBT:WTNT; MTAP:MTBT; WTAP:MTAP; and WTBT:MTBT.
I also summarized the occurrence of stress-response genes among sequences that were identified in the microarray experiment described by Tang et al. (2006) which was described in more detail in Chapter 2 of this dissertation.

The GOSSIP function of BLAST2GO 2.8 (https://www.blast2go.com/) was utilized to determine GO term enrichment for the entire set of differentially expressed genes in different comparisons (Blüthgen et al. 2005). I also used the online tool FungiFun2 (Priebe et al., 2014) to perform a functional annotation specifically of stress genes from the lists of differentially expressed genes using both the Illumina transcriptome and the microarray data sets (https://elbe.hki-jena.de/fungifun/).

4.3.5 Heatmaps

Gene expression patterns were visualized by creating heatmaps using log2 fold changes of genes generated from the transcriptome data. Those data were calculated as described in O’Connell et al. (2012). The expression ratio between the normalized counts of a gene in a developmental stage and the geometrical mean number of normalized reads across all the stages was calculated (Table 2.1 and 2.2). The log2FC is derived from this expression ratio and it was used to generate heatmaps of stress response gene expression profiles with the Genesis tool (Sturn et al. 2002).

4.3.6 Expression analysis of selected stress response genes in the WT under conditions of stress in vitro and in planta

4.3.6.1. In vitro analysis: The *C. graminicola* WT strain was cultured in 500 ml of Fries complete liquid medium (30 g sucrose, 5 g ammonium tartrate, 1.0 g ammonium nitrate, 1.0 g potassium phosphate, 0.48 g magnesium sulfate anhydrous, 1.0 g sodium chloride, 0.13 g calcium chloride, 1.0 g yeast extract/liter of H2O). Washed spores were added to produce a final concentration of 1x10^5 spores/ml, and the culture was incubated at 23°C on a rotary platform shaker at 15 rpm. After 5 days, the cultures were blended and 5 mls of the slurry was added to a new flask containing 50 ml of Fries minimal liquid medium and returned to the shaker. After 24 hours of recovery, the chemical treatments were added, and the mycelium was collected 12 hours later. The treatments
used and their final concentrations were tunicamycin (10 µg/ml) and Paraquat (3 mM). Mycelia were harvested under vacuum filtration and flash frozen in liquid nitrogen, then wrapped in aluminum foil packets and kept at -80°C until RNA extraction.

Appressoria of the WT were produced in vitro on polystyrene Petri dishes as described by Kleemann et al. (2008), with some modifications. *C. graminicola* spores were collected and washed three times, and 40 ml of a spore suspension at a concentration of 1 x 10^4 spores/ml was added to each Petri dish. Twenty hours later, each plate was inspected under the microscope to verify the presence of mature melanized appressoria. Trizol was added and appressoria were broken and scraped from the bottom using a sterile culture spreader. The slurry was recovered from 30 Petri plates in a total of 9ml of Trizol per replicate.

The RNA extraction was performed essentially as described in O’Connell et al. (2012), with a few modifications. Frozen mycelia were ground while still contained inside of the foil packet using a pestle. Around 100 mg of the powdered mycelia was added to a 2 ml Eppendorf tube with 1 µl of Trizol reagent (Invitrogen) for extraction. The cleanup step in the RNeasy Plant Mini Kit (Qiagen) was performed on the supernatant according to the manufacturer’s instructions, including the DNase A treatment.

For the first-strand cDNA synthesis, I used one µg of total RNA and the Superscript II reverse transcriptase kit (Invitrogen) with an oligodT primer. Semi-quantitative RT-PCRs were carried in 25 µl reactions and consisted of 0.1 µM of each primer, 0.2 mM each dNTP, 0.25 units of Taq DNA Polymerase (Life Technologies) and 1.5 nM MgCl₂. Thermal cycling was performed as follows: 94°C for 3 minutes followed by 30 cycles of amplification at 94°C for 45 s, 60°C for 30 sec and 72°C for 1 min. Actin (GLRG_03056) was used as an internal control. Sequential dilutions of cDNA were used as template, with the concentrations determined by using amplification of the control gene to normalize across samples, and diluting appropriately so that the control gene was in an exponential range (Choquer et al. 2003).
Three stress-associated genes were tested: GLRG_10629 (BiP); GLRG_02684 (HAC1); and GLRG_01327 (PDI) (Table 4.1). The primers were designed to produce an amplicon that spanned at least one intron, so that any DNA contamination of the RNA could be easily detected. Primer sequences all had an annealing temperature of 60°C, except for the control actin (GLRG_00649), for which the annealing temperature was 56°C.

4.3.6.2. In planta visualization of expression of the BiP homolog of C. graminicola. I used the pSITE vectors (Chakrabarty et al. 2007), modified by Gong et al. (2015) for the Gateway technology (Invitrogen) to produce a reporter construct to investigate expression of the BiP protein in C. graminicola. The upstream 433 bp of the yeast Kar2 (Bip) homolog (GLRG_10629) was introduced upstream of the RFP coding region (Figure 4.1, Figures AIII.3 and AIII.4 in Appendix III). Primers used to amplify the promoter region of GLRG_10629 included adaptors as described in the protocol by Gong et al. (2015). The GLRG_10629 gene is highly expressed in planta, in both MT and WT, at all stages of development. I introduced the clones into C. graminicola by Agrobacterium-mediated transformation (Flowers and Vaillancourt 2005). I recovered five independent transformants and single-spored them before use. Two of these transformants were observed with the Olympus FV1000 (Olympus America Inc., Melville, NY, USA) laser-scanning confocal microscope using 543 nm laser line. Response to in vitro stresses were evaluated by growing hyphae on sterile glass slides in a thin film of media, as described in Chapter 2 of this dissertation, amended with DTT (20 mM), tunicamycin (10 µg/ml), menadione (125 µM), and Paraquat (3 mM). Leaf sheath inoculations, as described in Chapter 2 of this dissertation, were also performed.

4.2.6.3 Visualization of the endomembrane system in vitro under stress. The endomembrane system of C. graminicola was labeled by transforming with the plasmid pAN56-1-sGFP-HDEL, containing GFP linked to an HDEL membrane anchor driven by a constitutive glucoamylase promoter (Vinck et al. 2005) as described in Chapter 2 (Figure A1.5).

Response to in vitro stresses were evaluated by growing the hyphae on sterile glass slides in a thin film of media, as described in Chapter 2 of this dissertation,
amended with DTT (20 mM), tunicamycin (10 μg/ml), menadione (125 μM), and Paraquat (3 mM).

Transformants were observed with the Olympus FV1000 (Olympus America Inc., Melville, NY, USA) laser-scanning confocal microscope using 543 nm laser line.
4.4 Results

4.4.1 The MT was more sensitive than the WT or MT-C strains to most stress-inducing chemicals and treatments

The MT was significantly more sensitive than the WT or MT-C strains to tunicamycin (secretion stress); Paraquat and menadione (oxidative stress); Calcofluor (cell wall stress); sorbitol and NaCl (osmotic stress); and low temperatures (Figure 4.2A-D, Table 4.2). The MT did not appear to be more sensitive to DTT (secretion stress) (Figure 4.2, Table 4.2); high temperatures (Figure 4.2E); nutrient limitation (Figure 4.2E); or high or low pH (Figure 4.2E).

4.4.2 About one-fifth of the *C. graminicola* genome encodes genes that are predicted to be involved in stress response

I compiled a list of 2730 putative stress response genes in *C. graminicola* based on similarity to known stress response genes in other fungi (Table 4.3). The genes were classified into secretion, osmotic, oxidative, and cell wall stress related categories (Guillemette et al. 2007; Mathioni et al. 2011; Nikolaou et al. 2009; Karányi et al. 2013; Jørgensen et al. 2009). An additional category, “other stresses”, included miscellaneous genes that matched the FSRD or other literature sources, and genes that were linked to general stress responses. Some genes occupied more than one category. The genes related to stress response were generally highly conserved. Considering only the top hit in the NCBI database (excluding *C. graminicola* itself), only 98 genes out of 2730 had less than 50% identity to their homolog in the model fungi. The average percent identity of the proteins in each of the four pathways to their homologs ranged from 88 to 90%.

4.4.3 *Colletotrichum* spp. encode homologs for a majority of the genes in the yeast secretory, oxidative, osmotic, and cell wall integrity stress response pathways.

The secretion, oxidative, osmotic, and cell wall integrity stress response signaling pathways in *S. cerevisiae* are very well characterized, but filamentous fungi do not always have homologs of all of the yeast genes in the pathways (Nikolaou et al. 2009). Five sequenced *Colletotrichum* species, including *C.
graminicola, had putative homologs for most, but not all, of the yeast proteins in these four pathways (Figure 4.3; Table 4.3). It was a general theme that the components that act as sensors and the transcription factors that regulate genes responsive to stress were the least highly conserved, whereas the components of the secondary messenger signaling pathways that connected them were more highly conserved.

Yeast has more redundancy in many of the stress response genes than the filamentous fungi. For example, the cyclic AMP protein kinase A Tpk1p, which functions in the oxidative stress response pathway, has two isoforms (Tpk2p and Tpk3p) in yeast, but there appears to be only one homologous gene in each of the Colletotrichum species. Similarly, in the secretion stress pathway, putative homologs for only 58 of 63 yeast genes were found in C. graminicola (Figure 4.3A; Table 4.3). Only nine out of the total of 129 genes potentially involved in secretion stress response in C. graminicola occur as more than one copy in the genome (Table 4.3). The increased redundancy in yeast could be related to a genome duplication that occurred in the yeast lineage some time in its evolutionary history (Wolfe and Shields, 1997).

Homologs of twenty-eight of the 33 yeast genes in the osmotic stress response pathway were present in C. graminicola (Figure 4.3B; Table 4.3). One gene, HOT1, was not found in C. graminicola, C. sublineola or C. higginsianum, but was present in C. orbiculare and C. gloeosporioides. This transcription factor in yeast is necessary for induction of the glycerol biosynthetic genes GPD1 (NAD-dependent glycerol-3-phosphate dehydrogenase) and GPP2 (glycerol-3-phosphate phosphatase) (Rep et al., 1999). It is possible that a different gene performs this function in C. graminicola, since C. graminicola does have homologs of GPD1 and GPP2.

C. graminicola and the other Colletotrichum species had putative homologs for all of the genes in the yeast oxidative stress response pathways except one, YBP1 (Figure 4.3C; Table 4.3). The Ybp1p oxidizes cysteine residues of the transcription factor Yap1p, which results in it relocating to the nucleus in response to stress. All five Colletotrichum species have a homolog of YAP1, although it is not very highly conserved with yeast in any of the Colletotrichum species. If this putative YAP1 homolog plays the same role in oxidative stress
response as the yeast protein, then it must be activated by some other mechanism in *Colletotrichum*.

*C. graminicola* had 21 of the 24 genes in the yeast cell wall integrity pathway. Three of the genes are shared with other pathways (Figure 4.3D; Table 4.3). One of the genes that *C. graminicola* didn’t share, *WSC3*, was found in its close relative *C. sublineola*, and also *C. gloeosporioides*, although the level of similarity was very low. The Wsc3p is a sensor that responds to heat shock and other stresses affecting wall integrity by activating the PKC-MPK1 signaling pathway and in yeast has two other paralogs, Wsc1 and Wsc2. *C. higginsianum* just has one paralog of this gene, while the others species have at least two. The components of that signaling pathway were conserved in *C. graminicola* and the other four species, so different paralogs must function in each *Colletotrichum*.

I assembled the homologs from *C. graminicola* into proposed osmotic, oxidative, and cell wall stress signaling pathways by using templates from Nikolaou et al. (2009) that are based on yeast (Figure 4.4B, Figure 4.4C, Figure 4.4D, Table 4.3). For the secretion stress pathway I used the model presented by Guillemette et al. (2007) instead, amended with additional information from Jørgensen et al. (2009) (Figure 4.4A). These two papers describe the secretion stress response pathway of *A. niger*. The pathway of the filamentous fungi, including *A. niger*, differs in some respects from the yeast pathway, and is likely to be a better model for filamentous fungi like *C. graminicola* (Miskei et al. 2009; Duran et al. 2010). The other pathways have not been as well studied in filamentous fungi, so yeast remains the best model for those. These models provided me with a framework for interpretation of the transcriptome data (below).

### 4.4.4 Presence of UPRE in stress genes

The unfolded protein response elements (UPRE) are cis acting elements found in the 5’ upstream region of genes, where transcription factors will bind to activate the UPR. In yeast, those elements were found to be important for appropriate activation of the UPR pathway (Yoshida et al., 1998). The transcription factor Hac1 binds to these UPRE and this leads to activation of
the genes (Fordyce et al. 2012). *Hac1* itself also has a UPRE, and a mutant with a defect in this region had increased sensitivity to secretion stress (Ogawa and Mori, 2004). In the upstream regions of the homologs of *Hac1* and *Sln1* in *C. graminicola* I found the canonical UPRE 5’-CAGcGTG-3’ (Fordyce et al., 2012; Ogawa and Mori, 2004) (Figure 4.5). In *Hog1* I identified the ER stress response element (ERSE) sequence CCAATN₈CCACG, which differs by only one nucleotide from the motif identified in yeast (Roy and Lee, 1999). I did not identify classic UPRE sequences in *Ire1* (GLRG_10691), *BIP* (GLRG_10629), *Tsa1* (GLRG_10121), *Wsc1* (GLRG_06481), MFS transporter (GLRG_06379), Catalase (GLRG_05821), glucose-repressible protein (grg – GLRG_03168), *RHO1* (GLRG_05224), *PTC2* (GLRG_04244), *PDI* (GLRG_01327), *MNT2* (GLRG_00793), *BAS3* (GLRG_00201), or Actin (GLRG_00649).

4.4.5 Genes involved in response to stress, and particularly to secretion stress, are highly expressed *in planta*.

There are 58 genes in the entire stress list that are part of the top 100 most expressed genes in at least one stage in the WT or MT. Three of the genes that I had mapped to *C. graminicola* secretion stress response pathways were also among the 100 most highly expressed genes *in planta* (Table 4.3). All three are components of the secretion stress response pathway, involved in protein folding. BiP, the ER chaperone that has an important function in activating the UPR, was found among the top 100 in every *in planta* treatment (WTAP, WTBT, WTNT, MTAP, and MTBT) (Table 4.3). Homologs of *PDI1* and *CLX1*, also included in the secretion stress pathway, were among the top 100 in MTAP and MTBT, and *PDI1* was also among the most highly expressed genes in WTAP and WTNT. *PDI1* is a protein disulfide isomerase that makes and breaks disulfide bonds between cysteine residues during folding. *CLX1* (or *CNE1*) is a calnexin ER chaperone protein involved in folding and stability of glycoproteins. *PDI1* and *BiP* were both induced during *in vitro* induced secretion stress in *Trichoderma reesei* (Pakula et al. 2003). The other characterized pathways had no genes that were included in the top 100 most highly expressed genes.

There were 18 stress-related genes that were included among the top 100 most highly expressed genes across every treatment (WTAP, WTBT, WTNT, MTAP
and MTBT) (Table 4.3). As mentioned above, one of these was BiP, involved in secretion stress response. The other genes were from the “other stresses” category, and included heat shock proteins involved in protein stabilization, ribosomal proteins and ubiquitin responsible for turnover of protein populations, a homolog of the Neurospora crassa cpc-2 gene, involved in response to nitrogen starvation (Müller et al., 1995), and genes encoding glyceraldehyde 3-phosphate dehydrogenase and other proteins potentially related to repair and protection of DNA (Takaoka et al. 2014) (Table 4.3).

I used FungiFun2 program to make a functional annotation of the 58 stress-associated genes among the top 100 most highly expressed genes in at least one developmental stage (Table 4.4). Out of those, 40 genes were annotated into categories. The most highly represented categories were involved in protein binding, folding and stabilization, and 11 genes involved in protein synthesis including a ribosome protein (GLRG_06907). The UPR pathway and peroxidase reaction (related to oxidative stress response) account for 9 genes. There is one gene (GLRG_02292), described as being involved in prevention of apoptosis, that was highly expressed during biotrophic and necrotrophic stages. Anti-apoptotic genes were implicated in full pathogenicity of the necrotroph Botrytis cinerea (Shlezinger et al. 2011).

Among these 58 highly expressed stress-related genes, 16 were found only on the top-100 lists in the MT strain (Table 4.3). Three of them were annotated as part of the UPR pathway, two as involved in protein folding and stabilization, and one in oxidative stress reaction. The other three genes annotated were categorized as involved in energy (Table 4.4).

4.4.6 Patterns of differentially expressed genes in planta suggest that stress responses are most active early during infection by both the WT and the MT.

It is know that global stress responses can result in down-regulation of genes involved in growth, RNA metabolism and protein synthesis, and up-regulation of others that are important for adapting to the stress (Nadal et al. 2011; Al-Sheikh et al. 2004). A majority of the most highly expressed genes, both in the WT and the MT during all phases of development in planta were involved in primary metabolism, growth, and signal transduction, with about 20% in each
case involved in stress response. Genes involved in antioxidant activities were among those that were overrepresented in WTAP relative to WTBT in the molecular function category, suggesting there might be a higher level of oxidative stress response during WTAP vs WTBT. The transition to WTBT correlated with enrichment in primary metabolism genes, suggesting an increase in primary metabolic activity during WTBT vs WTAP. Categories associated with stress response and PCD were overrepresented during WTBT compared with WTNT, suggesting higher levels of stress response during WTBT. These data suggest that stress responses occur throughout development, but that they are more active during early phases, when the host cells are alive and actively defending themselves. In comparisons of the MT and the WT, categories of genes involved in stress response were not enriched in either case, either during AP or BT, suggesting that transcriptional activity related to stress response was similar in the two strains.

4.4.7 Relatively few stress responsive genes are differentially expressed in planta.

Most of the C. graminicola putative stress response genes, including those that were located in the putative response pathways, were relatively poorly expressed in planta (Table 4.3, Figure 4.6A-D). Very few genes mapped to these known stress pathways were significantly differentially expressed (Table 4.3, and Figure 4.6). Out of the 2730 genes in the stress table, only 17% (490, almost all in the “other stress genes” category) were differentially expressed in at least one comparison (Table 4.3, Figure 4.6E). This supports the possibility that stress response is deployed at a relatively constant level during all phases of in planta growth, both by the WT and by the MT.

In the secretion stress pathway, only two genes were differentially expressed (Figure 4.6A). Homologs of GLRG_07837 (Mns1) and GLRG_00793 (Mnt2) are involved in protein glycosylation. Mns1 is more highly expressed during WTBT and WTNT, and Mnt2 is more highly expressed earlier, during WTAP. Mns1 is also involved in the ERAD (Delic et al. 2013), which might indicate that during later developmental stages, more proteins are being misfolded and subjected to degradation. In the microarray data, there are six genes that are part of the
secretion stress pathway that are increased in expression in biotrophic hyphae compared with cultured mycelium. These genes are mostly involved in protein folding and metabolism of energy reserves. Two genes are more highly expressed during in vitro conditions, homologs of the amino acid permease Hnm1 (GLRG_10760) and the Rud3 protein involved in structural organization of the Golgi (GLRG_08651).

The putative osmotic stress pathway in C. graminicola (Figure 4.6B) has only one differentially expressed gene, which is more highly expressed during WTNT (Table 4.3). The yeast homolog of GLRG_03441 (Sln1) is a transmembrane protein that functions as an osmosensor (Rodriguez-Pena et al., 2010). It is also important for cell wall integrity. The homolog of PTC2 (GLRG_04244) has >2 fold change in the LCD, indicating it’s highly expressed in planta versus in vitro. In yeast, PTC2 together with PTC1 and PTC3 negatively regulate the HOG pathway (Young et al., 2002). The homolog of Ssk1 (GLRG_04594) has a <3 fold change in the LCD, indicating that it is up-regulated in vitro versus in planta. Ssk1 activates HOG1 in yeast under conditions of extreme osmotic stress. The Sln1 sensor only functions in lower osmolarities (Posas et al. 1996).

The model for the oxidative stress pathway in C. graminicola (Figure 4.6C) has two differentially expressed genes, Sln1 and Tsa1 (GLRG_03441). The first one is also present in the osmotic pathway. The second, Tsa1, encodes a peroxiredoxin, an important antioxidant in yeast that binds to YAP1 and participates in ROS detoxification. Yeast Tsa1 mutants are hypersensitive to secretion stress and to ROS (Weids and Grant, 2014; Wong et al., 2002). GLRG_03441 is less highly expressed in the C. graminicola MT versus the WT, during both AP and BT.

Only one of the genes present in the cell wall integrity pathway (Figure 4.6D), is differentially expressed. GLRG_06481 is a homolog of Wsc1, a plasma-membrane sensor that responds to changes in the cell wall and activates Rom, the first step in the activation of the MAP kinase cascade (Igual and Estruch, 2000; Nikolaou et al., 2009). This gene is more highly expressed during WTNT than during earlier stages of development. According to the LCD, the homolog of Sec3 is more highly expressed in planta versus in vitro. Sec3 functions in
transport of secretory vesicles from the Golgi to the plasma membrane, in a pathway dependent on the Rho protein but independent of the normal secretory pathway or the actin cytoskeleton (Levin 2005).

For the "other stress genes" category there also appeared to be groups of genes that were more highly expressed either during AP or during NT (Figure 4.6E; Table 4.5). Out of 2,543 genes, 491 of them had a >2-fold change (log2) in at least one condition. More than half (214) of the annotated genes were categorized by FunCat as primarily involved in metabolism, followed by 76 (19%) involved in cellular transport. There were 69 genes that were classified as cell cycle, DNA processing, cell rescue, defense and virulence, accounting for 17.5 % of the annotated genes, identified in Figure 4.6E by the name stress related. These included stress response [19 genes including catalases (2), Hsp70 (2) and superoxide dismutase (1)]; DNA repair (16 genes); detoxification by export (13 genes); and other classes with fewer genes like nutrient starvation response (5), and pH stress response (3).

I examined a microarray dataset produced by Tang et al (2006), representing differentially expressed transcripts from biotrophic hyphae captured by laser capture and compared with hyphae that were cultured in vitro. These authors were kind enough to share these unpublished data with me (Table 4.3 - LCD column). There are 797 stress-related genes represented in this dataset. Among those, 127 were more highly expressed in planta (logFC > 1), and 136 were more highly expressed in vitro (logFC of < -1). In the secretion stress pathway, ten genes were more highly expressed in planta and three were higher in vitro (Table 4.3, Figure 4.4A: genes present in the pathway are marked in pink for higher during in planta and yellow during in vitro). The osmotic pathway included one gene that was higher in planta (Ptc2) and one that was lower (Ssk1) (Figure 4.4B). The oxidative pathway has two genes that were higher in vitro: one is Ssk1, shared with osmotic pathway, and the other is Bcy1 (Figure 4.4C). In the cell wall integrity pathway, only one gene, the exocyst complex gene Sec3, is differentially regulated, being more highly expressed in planta (Figures 4.4D).

Among the 127 genes with higher expression in planta, 57 could be assigned to 12 different categories by using FunCat (Figure 4.7A). These included
categories relevant to UPR, DNA damage response and DNA repair, protein folding, and general stress response, indicating that the biotrophic hyphae were responding to these stresses more actively than hyphae in culture. Among the 136 genes with higher \textit{in vitro} expression, only 14 could be classified by FunCat, and all were either in detoxification or drug/toxin transport categories (Figure 4.7B). These 14 genes were predicted to encode either major facilitator superfamily transporters or ABC-2 type transporters.

\textbf{4.4.8 Expression of selected stress response genes \textit{in vitro} in response to chemically induced stress.}

Selected genes were tested by semi-quantitative RT-PCR in response to chemical stress (Figure 4.8). The housekeeping gene actin was used as the RT-PCR normalization control. I used the same amount of total RNA to make cDNA, but each \textit{in planta} treatment had different amounts of fungal biomass (Table 2.3). Unfortunately the dilutions that I tried did not give consistently good results, and so it is difficult to quantify the expression of the genes relative to one another in every treatment. I can tell that the stress genes I selected (\textit{BiP}, \textit{Hac1} and \textit{PDI}) are being expressed in all the tissues. In the \textit{in vitro} appressoria (IVAP) HAC1 seems to be expressed at a high level while BiP and PDI seem to have lower expression. In yeast, HAC1 is constitutively expressed in all tissues, being regulated post-transcriptionally by differential splicing, and transcript levels vary relatively little (Schröder et al., 2003). I found no evidence for differential splicing of the putative \textit{C. graminicola} HAC1 homolog. Both BiP and PDI seem to be more highly expressed \textit{in planta} AP versus IVAP. HAC1 also seems to be expressed at relatively high levels \textit{in planta} compared with \textit{in vitro} (Fries medium). All three genes seem to be expressed at lower levels during BT versus in AP and NT. Because of the issues with dilutions, it was not possible for me to tell whether treatment with chemical inducers of stress (Paraquat and tunicamycin) induced the expression of any of these three genes. These experiments must be repeated before I can make any firm conclusions about the identities of these three genes and their roles in stress response in \textit{C. graminicola}. 
4.4.9 Stress altered the structure of the MT Cpr1 transcripts in vitro compared with in planta

RT-PCR revealed the presence of the same transcript variants of Cpr1 in the MT in vitro, in the presence or absence of stress, as were previously observed in planta and in vitro in Chapter 2 (Figure 4.9). Sequencing results of several different clones confirmed that only the normal transcript was detectable in the WT, either with or without stress (Figure 4.10A). In the mutant, sequencing revealed the same intron retention variants that were identified previously (in Chapter 2) in all treatments, including the control without stress. Interestingly, a novel pattern of intron splicing was identified in the MT treated with tunicamycin (Figure 4.10B). An aberrant intron is spliced from positions 391-CG^GTTGGG to ATGGCAG^CA-445, whereas the normal intron is from 446-CG^GTAAGA to ACTTCGCAG^TG-505. There are no conserved splice motifs at the junctions of this novel intron in the “plus” strand, but there are in the “minus” strand.

4.4.10 In planta visualization of the BiP stress response protein

Fluorescent reporter constructs with the C. graminicola BiP chaperone were used as another approach to visualize stress response in planta. The transcriptome results show BiP as one of the most highly-expressed stress genes in all stages of development in planta. Hyphae of transformants containing reporter constructs in which RFP was linked to the promoter of the BiP gene were treated in vitro with chemicals known to induce stress. RFP accumulated in hyphae and appressoria in response to DTT, tunicamycin, Paraquat, and especially menadione (Figure 4.11). There was no RFP visible in the untreated controls. I could detect strong red fluorescence in spores, and particularly in primary hyphae in planta. These results suggest that BiP expression is induced in response to stress in vitro and in planta.

4.4.11 In vitro response of endomembrane system to stress chemicals

I used a WT-HDEL transformed isolate to visualize the fungal endomembrane system, already described in Chapter 2, during in vitro stress. Analysis of the transformant grown in minimal Fries medium amended with the chemical
tunicamycin show an increase of fluorescence when compared to the control sample, grown only in minimal Fries liquid medium (Figure 4.12). It is important to point out that all treatments were analyzed at the same exposure. The hyphae grown in the presence of stress became swollen, and produced multiple branches. The appearance was quite reminiscent of the growth of primary hyphae like those that are found inside the plant (Panel C and D). At 1000x magnification (Panel E) it is possible to clearly identify the ER membrane pattern surrounding the nuclei.
4.5 Discussion

In this chapter, I explored the hypothesis that defects in pathogenicity displayed by the *C. graminicola cpr1* MT were related to deficiencies in stress response. This hypothesis was prompted by reports that the *Cpr1* homolog in some other fungi is induced during *in vitro* conditions that result in secretion or nutritional stress (Guillemette et al. 2007; Jørgensen et al. 2009). It is assumed that the living plant imposes a stressful environment on the pathogen, and if the MT is unable to adapt to stress, this could result in its being unable to establish a successful infection. For example, mutants of *M. oryzae* deficient in oxidative stress response were non-pathogenic to rice (Guo et al. 2011).

My first prediction based on this hypothesis was that the MT would be more sensitive than the WT to chemically-induced stresses *in vitro*. The *cpr1* MT strain has no apparent differences from the WT strain *in vitro*, other than a somewhat slower rate of radial growth (Thon et al. 2000; Thon et al. 2002; Torres et al. 2013; Venard and Vaillancourt 2007). My experiments showed that the mutant is more sensitive than the WT to a range of stress-inducing chemicals, including compounds reported to induce secretion stress, oxidative stress, and especially cell wall stress (Figures 4.2). The mutant was also more sensitive to cold temperatures, but its sensitivity to heat, and to nutritional and pH stress, appeared to be unaltered. There is a lot of cross-talk that occurs between stress pathways, so this may explain why the MT appears to have a relatively broad deficiency in stress response.

To induce secretion stress, I used two different chemicals: tunicamycin, and DTT (Guillemette et al. 2007). The MT was sensitive to tunicamycin but not DTT. DTT not only causes secretion stress, it is also an antioxidant (Liu et al. 1999). Secretion stress is known to cause the accumulation of ROS and thus, oxidative stress (Haynes et al., 2004). So one possibility is that DTT, acting as an antioxidant, modulates the toxicity of the secretion stress that it presumably induces in *C. graminicola*. It is important to point out that, even though all of the chemicals I used had negative effects on the growth and development of *C. graminicola*, I do not have direct evidence that they were inducing the stress pathways that have been linked to them in the literature. It will be important to investigate the expression of known stress pathway genes, to confirm that they
are induced by these chemicals as expected. Unfortunately I was not able to accomplish this, although I did develop some tools for semi-quantitative PCR analysis that would help to answer this question. Northern blots would be the best type of experiment to do, at some point.

My in silico analysis of the C. graminicola genome suggested that 22% of the predicted proteins are stress-related. I developed detailed proposals for several putative stress response pathways in C. graminicola, based on comparisons with the literature, particularly reports related to S. cerevisiae. Most of the references I found describing stress responses in filamentous fungi used yeast as a model, even though there seem to be some differences between pathways in yeast and filamentous fungi. For example, A. nidulans requires two genes, Hog1 and Pbs2, to activate osmotic stress response, whereas yeast only requires Hog1 (Furukawa et al., 2005). Target genes of the pathways were also found to differ in some cases from those in yeast, as shown in this paper about cell wall stress signaling (Fujioka et al. 2007).

I was able to organize stress gene homologs into putative pathways involved in secretion, osmotic, oxidative, and cell wall stress. The proteins that are key regulators in the pathways were the most highly conserved between yeast and Colletotrichum (Figure 4.3). For example, BIP in the secretion stress pathway, Hog1 in the osmotic stress and oxidative stress pathways, and Rho1 in the cell wall stress pathway had greater than 70% sequence identity with the yeast proteins. Nothing had been done with stress response before in C. graminicola, so this work will be valuable for developing hypotheses for future studies related to stress.

Despite the high level of conservation, I was not able to find evidence for conservation of promoter elements that are present in stress genes and regulate their expression in yeast. The promoter region of the gene in the 5'UTR of some stress genes show what is called ER stress response elements (ERSE) or Unfolded Protein Response Elements (UPRE). ERSE, with a consensus of CCAATN=CCACG, is suggested to be a sequence that coregulate stress genes during stress conditions by a common factor, activating different genes involved in the Unfolded Protein Response (UPR). In yeast, it is interesting to notice that while ERSE leads to a increase in translation and protein synthesis of genes
involved in stress response, expression of genes involved in protein synthesis are actually reduced, maybe in an effort to protect the cell while trying to adapt to the new environment (Berry and Gasch, 2008). Although this activates some genes, it is a mild response in preparation for what type of stress might actually be happening. Then other stress genes will be activated, involved in each type of stress. In *C. graminicola*, I found those elements in the upstream region of three out of 16 genes involved in stress response that I checked (Figure 4.5), although all of them have parts of the ERSE sequence, but not the complete sequence. Only *Hog1* in *C. graminicola* has an intact ERSE. *Sln1* has a UPRE sequence in the promoter region, being two components of the osmotic stress pathway. The other gene is *Hac1*, a transcription factor important in the secretion stress and UPR. It could be that in *C. graminicola* activation of ERSE doesn’t depend on the entire consensus element sequence from yeast or that regulation is not controlled by that at all. The presence of these elements in the entire set of putative stress-related genes needs further study.

My second prediction for this chapter was that WT and MT strains would express stress response genes *in planta*, indicating that they were experiencing, and reacting to, stress. All of the genes that I included in the stress response pathways were expressed *in planta*. Although there were a few exceptions (e.g. the gene encoding the BiP homolog), most of the genes were not expressed at very high levels. Signaling genes are frequently not highly expressed, and the expression of stress genes is sometimes transient (López-Maury et al. 2008). Both of these factors could account for the generally low expression levels. Alternatively, the stress response genes might be regulated post-transcriptionally (Lackner and Bähler, 2008). The paper by Guilllemette et al. (2007) found a transcriptional increase of only 2-fold for the Cpr1 homolog, but they reported that there was a 7-fold increase in translation efficiency during secretion stress conditions, due to differential polysome loading.

According to my interpretation of the laser capture microarray data of Tang et al. (2006), numerous stress genes were induced *in planta* versus *in vitro*, suggesting that the biotrophic hyphae are experiencing, and responding to, greater levels of stress. In the paper by Vargas et al. (2012), it was reported that the maize plant expresses defense genes at a high level during biotrophy,
and ROS production during biotrophy has also been demonstrated by cytological assays (Vargas et al. 2012; Torres et al. 2013).

It has been suggested that the biotrophic primary hyphae of hemibiotrophs like *C. graminicola* could be good models for the haustoria of obligate biotrophs like rusts and powdery mildews, which cannot be cultured. It has generally been believed that obligate biotrophs secrete effectors that suppress host defenses, and thus they don’t need to express stress response pathways, in contrast to necrotrophs that actually induce plant cell death PCD (Glazebrook, 2005; Schulze-Lefert and Panstruga, 2003). But several recent papers are changing this idea. For example, work in *Ustilago maydis* found that to successfully infect the plant the pathogen must induce antioxidant pathways (Doehlemann et al. 2008). Proteomic studies identified several heat shock proteins expressed in the haustoria of powdery mildew (L V Bindschedler et al. 2009). Several transcriptome studies of isolated haustoria of rust and powdery mildew fungi also contain evidence for activity of various stress response pathways (Garnica et al. 2013; Weßling et al. 2012; Link et al. 2014).

My third prediction was that the MT would differ from the WT in the expression of stress response genes *in planta* and *in vitro* under stress conditions. However, very few of the stress response genes I identified were differentially expressed in the transcriptome data, either between MT and WT, or between different stages of development in the WT. Stress response genes can be regulated post-transcriptionally (Schröder and Kaufman, 2005), so the lack of differences could mean that the regulation of the genes is primarily post-transcriptional. Another possibility is that the *in planta* environment is uniformly stressful. The up-regulation of many genes in the LCD microarray data supports the idea that even the biotrophic hyphae are experiencing significant levels of stress.

Characterization of the induced stress response genes from the LCD microarray data suggests that *C. graminicola* biotrophic hyphae are responding to secretion stress by activation of the UPR response. There is also evidence in the transcriptome data for activity of the UPR during all developmental stages (Figure 4.7, Table 4.4, Table 4.5). It is expected that *C. graminicola* would need to secrete a variety of different proteins to successfully infect, colonize, and
finally rot the host tissues, and this high requirement for secretory activity could
certainly lead to secretion stress. So it is not surprising to find UPR activity.
Apparently the MT can express these genes, and therefore an inability to
transcribe the genes is not the reason for its increased sensitivity to
tunicamycin, or for its non-pathogenic phenotype.

Preliminary data from my semi-quantitative RT-PCR experiments (Figure 4.8)
supports the idea that UPR-associated stress genes are induced in planta
during all developmental phases. There seemed to be less expression in vitro,
in appressoria produced on Petri plates. Unfortunately I could not confirm that
tunicamycin and menadione induced expression of these genes in vitro using
this technique. However, both chemicals, as well as DTT and Paraquat,
induced expression of an RFP reporter linked to the BiP promoter (Figure 4.11).
Furthermore, the reporter was also strongly expressed in biotrophic hyphae in planta. One effect of the induction of the UPR is an increase in secretory
capacity, which can be visualized as an increase in the volume of the
endomembrane system in yeast (Bernales et al., 2006). Visualization of the
GFP_HDEL endomembrane system in the WT strain treated with tunicamycin
seemed to show an increase in fluorescence when compared to the control
(Figure 4.12).

As I described in Chapter 2 of this dissertation, I found that the MT produced
several variant cpr1 transcripts in vitro and in planta, while the WT produced
only the predicted transcript (Figure 4.9). I wanted to see if the same variants
were produced in the MT or WT exposed to stress. I observed that only the MT
produced variants, which appeared to be similar in size and number to those
that I had observed in Chapter 2 (Figure 4.10). However, cloning of the
amplicons revealed a new variant that I had not seen before, which had a novel
intron removed that was just upstream of the normal intron, which was retained
(Figure 4.10B). There were relatively few reads that matched any of the splice
variants, so I can’t say for sure that this variant is specific to the tunicamycin
treatment. Further studies are needed to confirm the frequencies of the different
splice variants that I saw, and whether any are specific to any particular
condition. I am currently doing an Illumina sequencing experiment to try to
address this.
Interestingly, the putative intron in the tunicamycin-associated variant did not have canonical splice signals on the positive strand, but it did have them on the minus strand, suggesting that these reads could be from a transcript that was being produced from the opposite strand. None of the gene prediction programs I used predicted any ORFs on the minus strand, but that doesn’t necessarily mean there isn’t a gene there. Although I have been assuming that the stress response and pathogenicity phenotypes of the MT are linked, it’s also possible that the stress response phenotype is due to a deficiency in an unknown stress response gene on the minus strand, and not Cpr1 itself. Unfortunately I was not able to determine which strand the transcriptome reads originated from, so this question remains unanswered for now.

In summary, the work I did for this chapter indicated that the MT and WT are both experiencing and responding to stress, particularly secretion stress, \textit{in planta}; that there is no difference between them at the transcriptional level; and that stress appears to be induced \textit{in planta} in comparison with \textit{in vitro}, equally across all developmental phases. In addition to these findings, I produced descriptive models of major secretory stress pathways in \textit{C. graminicola}, and fungal strains that can potentially be used to monitor stress response \textit{in planta} in future studies.
Figure 4. Illustration of the pFPL Gateway vector construct used to create protein fusions between *C. graminicola* BiP promoter region and the red fluorescent protein reporter gene. Figure kindly provided by Dr. Mark Farman and modified to show where the *BiP* promoter was added.
Figure 4. Growth of *C. graminicola* strains during chemical stress. A) Sensitivity of wild-type (WT), mutant (MT) and complement (MT-C) strains to secretion stress chemicals Tunicamycin and DTT. Values are expressed as cm (left), or percentage growth in the absence of the chemical (%) (right).
Figure 4.2 Growth of *C. graminicola* strains during chemical stress. B) Sensitivity of wild-type (WT), mutant (MUT) and complement (C) strains to oxidative stress chemicals Paraquat and Menadione. Values are expressed as cm (left), or percentage growth in the absence of the chemical (%) (right).
Figure 4.2 Growth of *C. graminicola* strains during chemical stress. Sensitivity of wild-type (WT), mutant (MUT) and complement (C) strains to osmotic stress chemicals Sorbitol and NaCl. Values are expressed as cm (left), or percentage growth in the absence of the chemical (%) (right).
Figure 4.2 Growth of *C. graminicola* strains during chemical stress. D) Sensitivity of wild-type (WT), mutant (MUT) and complement (C) strains to cell wall stress chemical Calcofluor. Values are expressed as cm (left), or percentage growth in the absence of the chemical (%) (right).
Figure 4.2 Growth of *C. graminicola* strains during chemical stress. E) Sensitivity of wild-type (WT), mutant (MUT) and complement (C) strains to temperatures. Sensitivity of wild-type (WT), mutant (MUT) and complement (C) strains to nutritional and pH changes. Values are expressed as cm (left), or percentage growth in the absence of the chemical (%) (right).
Figure 4.1: Stress pathway models. A model of the secretion stress pathway together with mean normalized read counts for each gene across the following stages: wild-type appressoria (WTAP), wild-type biotrophic (WTB), wild-type necrotrophic (WNT), and mutant appressoria (MTAP) and mutant biotrophic (MTB). Reads in red mean that the gene was differentially expressed between biotrophic hyphae in the LCD. Genes in blue mean that they are unique to that pathway, pink means that they are shared with other pathways.
FIGURE 4.4 Stress pathway models. B) Model of the osmotic stress pathway together with mean normalized read counts for each gene across the following stages: WTAP, WTBT, WTNT, MTAP and MTBT. Reads in red mean that the gene was differentially expressed between different stages. Genes in blue mean that they are unique to that pathway, pink means that they are shared with other pathways.
FIGURE 4.4 Stress pathway models. C) Model of the oxidative stress pathway together with mean normalized read counts for each gene across the following stages: WTAP, WTBT, WTNT, MTAP and MTBT. Reads in red mean that the gene was differentially expressed between different stages. Genes in blue mean that they are unique to that pathway, pink means that they are shared with other pathways.
FIGURE 4.4 Stress pathway models. D) Model of the cell wall stress pathway together with mean normalized read counts for each gene across the following stages: WTAP, WTBT, WTNT, MTAP and MTBT. Reads in red mean that the gene was differentially expressed between different stages. Genes in blue mean that they are unique to that pathway, pink means that they are shared with other pathways.
Figure 4.5 Placement of response elements in the 5' upstream region of genes involved in stress. UPRE: unfolded protein response elements; ERSE: ER stress response element.
Figure 4.6 Heat maps. A) Secretion stress pathway, B) Osmotic stress pathway, C) Oxidative stress pathway, D) cell wall stress pathway.
FIGURE 4.6 Heat maps. E) Heat maps of gene expression of those categorized as "Other stresses" (Table 4.3) with a two-fold log2 in at least one condition. Genes were annotated in different categories by FungiFun2.
Figure 4.7 Analysis of stress-related genes present in the laser capture data with a logFC higher than 1. A) Genes with higher expression \textit{in vitro} and B) genes with higher expression \textit{in planta}.
Figure 4.8 RT-PCR with BiP, Hac and PDI during *in vitro* and *in planta* conditions. Actin (Act) is used as a control. IVAP: *in vitro* appressoria, AP: appressoria, BT: biotrophic, NT: necrotrophic, Fries: minimal Fries medium, PQ: Paraquat, TUN: tunicamycin. The numbers on top of the lanes represent the ratio between that lane and the control actin, as calculated by the program GelQuantNET (BiochemLabSolutions).
Figure 4.9 Effect of stress on Cpr1 transcripts *in vitro*. RT-PCR amplification of material grown *in planta* and *in vitro* from WT and MT strains. Treatments are *in vitro* appressoria (IVAP), appressoria (AP), biotrophic (BT) and necrotrophic (NT), as well as *in vitro* fries control (Fries), Paraquat (PQ) and tunicamycin (TUN). Ladder (L) is 1 kb Plus DNA ladder from Life Technologies. Primers used CPR1intF4xCPR1intR4.
Figure 4.10 Transcript sequencing results of *Cpr1* gene in the WT and MT strains. A) While WT maintains the predicted intron in all the conditions tested, the MT shows intron retention and alternative splicing, as well as the normal intron. B) Close up of the three possible intro variants of *Cpr1* gene. The first one is the predicted intron, here represented by a sequencing of WT during Paraquat treatment. The second line shows the intron retention, in this example shown by MT control on minimal Fries medium. And last, the third version of the intron found only in MT during tunicamycin treatment.
Figure 4.11 Expression of BiP (GLRG_10629) chaperone reporter *in vitro* and *in planta*. All pictures were taken in the confocal microscope at 520 V/ 543 nm. Magnification of 400x or 1200x.
Figure 4.12 WT-HDEL tagged isolate visualized in the confocal microscope at 520 V/ 543 nm during *in vitro* growth with the chemical tunicamycin.
Chapter 5

Concluding remarks

I have focused in my dissertation research on a very interesting non-pathogenic mutant of the maize anthracnose fungus *Colletotrichum graminicola*. This mutant was produced in our laboratory by insertional mutagenesis quite a few years ago, and it has been the subject of study by various laboratory members ever since (Mims and Vaillancourt, 2002; Thon et al., 2002, 2000; Torres et al., 2013; Venard and Vaillancourt, 2007a). The thing that is very interesting about this mutant is that it is conditional: it grows normally in culture; and it germinates, produces appressoria, and penetrates the host normally. It specifically fails to establish biotrophic hyphae, and thus to produce a successful infection. The previous work in the laboratory has taught us much, but so far it has been unable to explain the conditional nature of this mutation.

My arrival in the Vaillancourt laboratory coincided with a major effort to obtain new genomic resources for *C. graminicola*, and to apply those to understanding its pathogenicity to maize. A major part of my work has been on generating and analyzing the genome and transcriptome of *C. graminicola*. I first came to the Vaillancourt lab for a short research sabbatical right after I finished my undergraduate degree. While I was here, I helped to extract the DNA of M1.001 (aka WT) that was sent for sequencing at the Broad Institute. Little did I know that I would come back and work so intensively with this strain and this genome! When I joined the lab as a PhD student, one of the first things I did was to extract DNA from *C. graminicola* M5.001 and from *C. sublineola* CgSi1, both of which were sequenced by the AGTC and also used in my analysis. Later, I prepared RNA samples for the transcriptome study, together with my fellow graduate student at the time, Dr. Maria Torres. I have focused a lot of my time and effort on organizing, summarizing, and analyzing all of these different datasets. In the process of doing this, I have developed various resources that will make it easier for future researchers to access and interpret these data.
In spite of the fact that the mutant had been in the lab for more than ten years, nobody had ever done a comprehensive analysis of the insertion site in the MT Cpr1 allele. The genome made this an easier task (kind of!), and so I used a combination of PCR amplification, sequencing, and Southern blotting to characterize the mutation in detail. Mapping of the transcriptome reads to the map of the mutant allele revealed that intron splicing appeared to be altered in the MT, compared with the WT. I confirmed by RT-PCR, cloning, and sequencing, the presence of at least three variant splice forms in the MT that don’t seem to occur in the WT. To further characterize this phenomenon, I prepared RNA from 60 samples, representing MT, WT, and MT-C, in planta, and in vitro in the presence or absence of stress, for high-throughput Illumina sequencing. These samples are currently being processed by the AGTC.

The insertion in the MT was known to be in the 3’UTR, 19 bp downstream from the stop codon of the Cpr1 ORF. This certainly implied that the 3’UTR sequence of the MT and WT would be different, but nothing was known about the precise nature of the MT 3’UTR length or sequence. I found several different “nested” versions of the 3’UTR for both strains in planta by using a PCR protocol to amplify the poly(A) regions. A former postdoc in the lab had used RACE to characterize the 3’UTR of the WT in vitro, and she found only a single version, which was a bit longer than the one predicted in Thon et al., (2002). None of the versions I cloned looked exactly like hers. One possibility is that my experiment was flawed: additional methods should be applied to characterizing the 3’UTRs of both strains in planta to confirm my results. Another, more interesting possibility is that the WT 3’UTR varies in planta, and that these variations have some significance in function. The MT 3’UTR differs from the WT in both length and sequence, and this could affect those functions. For example, the 3’UTR can regulate transcript stability, transcript localization and translation efficiency, and intron splicing, among other things. It is possible that the variation in intron splicing in the MT is a result of the altered 3’UTR sequence.

The CPR1 protein is predicted to comprise part of the signal peptidase complex, which is the first step in the canonical eukaryotic secretory pathway. I used the genome and transcriptome data to undertake comprehensive analyses of the
putative secretory pathway of *C. graminicola*. I also developed a WT strain expressing a RFP-CPR1 chimeric protein, and developed I developed some transformants that expressed an HDEL-GFP anchored in the ER membrane to visualize the endomembrane system in living cells of the WT, MT, and MT-C strains. I did not observe any differences between the WT and MT in the apparent expression or activity of the secretory pathway. Nonetheless, the tools that I have developed will help for future studies designed to test the hypothesis that secretion activity varies in the MT vs WT strain.

One hypothesis to explain why the MT is nonpathogenic is that it fails to secrete necessary effector proteins. I did not address this hypothesis directly, but I undertook a comprehensive “*in silico*” comparative genomic and transcriptomic analysis to characterize the effectorome of *C. graminicola*, and to identify the most likely candidates for effectors that could be involved in the establishment of biotrophy. My analysis is already being applied by a visiting scientist in our laboratory, who is investigating polymorphisms in these effectors among different strains of *C. graminicola*. I am confident that the tools and the data I developed will continue to facilitate future research in our laboratory on the mechanisms of pathogenicity in *C. graminicola*.

Even though it’s undeniable that the signal peptidase complex is critical for secretion, the idea that all secreted effectors have a signal peptide and are secreted thru the ER-Golgi canonical secretory pathway has recently been challenged in fungi. A recent study on yeast showed that UPR can trigger an unconventional secretion pathway for misfolded and excessive proteins to be delivered into the extracellular space (Miller et al., 2010). Several studies in human fungal pathogens show that two thirds of the proteins known to be secreted by them are exported via alternative pathways (reviewed by Rodrigues et al. 2013). A recent study with the plant pathogens *Phytophthora sojae* and *Verticillium dahliae* showed a protein that affected salycilic acid response pathway in plants and that did not have the canonical signal peptide even though they were translocated to the plant cytoplasm (Liu et al. 2014). In *M. oryzae* Giraldo et al., (2013) found that cytoplasmic and apoplastic effectors are secreted by two different secretion pathways: the exocyst complex and the conventional ER-Golgi secretion pathway, respectively (Giraldo et al. 2013).
Thus, it is important to keep an open mind about the possible function of the CPR1 protein, and think beyond a possible role in secretion.

Reports from work with other fungi suggested that the CPR1 protein could be involved in adaptation to secretion stress. To test this possibility, I characterized the reaction of the mutant to different stresses \textit{in vitro}. The race tube test that I developed to test the sensitivity of the fungi to stress-inducing chemicals is already being used by other student in the lab to characterize the reaction of other strains to fungicides. I used the genome and transcriptome data to develop models for the major stress pathways in \textit{C. graminicola}. To my knowledge stress response has not been characterized in \textit{Colletotrichum} previously. My analyses did not reveal any major differences in the expression of stress response genes \textit{in planta} between the MT and WT, or across the various WT developmental stages \textit{in planta}. On the other hand, the laser-capture microdissection dataset that I analyzed strongly suggested that stress genes were induced \textit{in planta} in comparison to culture, supporting the idea that the plant is a stressful environment. My analysis using a WT strain transformed with a BiP-RFP reporter construct also suggested that hyphae are reacting to stress when growing inside the plant. This strain may provide a useful tool for monitoring stress response in the living host-pathogen interaction.

My data generally support the idea that post-transcriptional differences play an important role in gene regulation in the \textit{C. graminicola}-maize interaction, and future work should be focused at the protein level. There have been relatively few proteomics studies for pathogens during infection of plants, partly because there are many technical difficulties in obtaining and identifying the full range of proteins from the interaction. There have been a few proteomics studies that have implicated stress response (Bindschedler et al., 2009), and the production of small secreted proteins (Rep et al., 2004) in pathogenicity. Such studies will certainly become more accessible, and more common, in the future. It is important to emphasize that the genomic, bioinformatics, and transcriptomics analyses I’ve done for this dissertation will play an essential role in designing and interpreting future protein and proteomics studies of the \textit{C. graminicola} WT and MT.
Appendix I
Supplemental Material for Chapter 2

Verification of Transcriptome Sequencing Results

AI.1 Background

Dr. Maria Torres and I prepared the RNA samples that were sent to the Texas AgriLife Genomes and Bioinformatics Service Center for transcriptome sequencing. We inoculated leaf sheaths with two *C. graminicola* isolates, the M.1001 wild-type (WT) and the *cpr1* mutant (MT). For the WT strain, we collected appressoria (WTAP), biotrophic (WTBT), and necrotrophic (WTNT) developmental stages, and for the MT we collected appressoria (MTAP), and biotrophic (MTBT) stages. A summary of the WT transcriptome data was published in O’Connell et al. (2012).

The MT has an insertion in the 3’UTR of *Cpr1* (*Colletotrichum* pathogenicity related gene 1), a homolog of the *Spc3* gene in yeast. The MT is unique appears to produce a very small quantity of normal transcript, which however is sufficient for normal growth *in vitro* (Thon et al. 2002). 3’UTR sequences are implicated in regulating transcript cleavage and polyadenylation, controlling alternative polyadenylation, nuclear export, transcript stability, translation efficiency, and mRNA targeting (Grzybowska, Wilczynska, and Siedlecki 2001). The insertion in the 3’UTR of *Cpr1* might explain the reduced transcript levels in the MT.

In my work, described in Chapter 2, to analyze the MT *Cpr1* transcript, I wanted to A) discover if intron splicing in the MT differs from the WT; and B) characterize the sequence of the MT 3’UTR. I was given the sequence of the plasmid pCB1636 that was used in the Restriction-Enzyme Mediated Integration (REMI) mutation by Dr. Jim Sweigard, who created the plasmid, so I could identify transcripts from the insertion plasmid in the transcriptome, if they existed. The work by Thon and collaborators (Thon et al. 2002) described the mutant as containing 1.5 copies of the plasmid inserted 19 bp downstream of
the stop codon in the Cpr1 gene. I used this to create a hypothetical map of the insertion (Figure AI.1. Mutant). After that, I used BLAST to find reads in the transcriptome data that matched. In the process of doing this mapping, I discovered an error in our transcriptome data, which had probably been caused by a sample mix-up at Texas Agrilife. Below I will describe how I found out about this error and what we did about it.

AI.2 Material and Methods

Sample preparation and RNA extraction are detailed in the supplemental data in O’Connell et al. (2012) and in Chapter 2 of this dissertation. In short, maize leaf sheaths were inoculated with 5x10⁵ spores/ml of C. graminicola, either the wild-type or the mutant strain. The sheaths were collected at appressoria stage, intracellular biotrophic hyphae and necrotrophic stages with secondary hyphae visible. Each leaf sheath was confirmed with a light microscope and they were trimmed to include only the inoculated area. Appressoria and biotrophic leaf sheaths were also shaved to remove uninfected cells, trying to increase our fungal reads. The sheath pieces were maintained at -80°C and RNA was extracted by combining Trizol reagent (Invitrogen) and purification protocol from RNeasy Plant Mini Kit (Qiagen) with DNase A digestion. RNA integrity and quantity were measure with an Agilent 2010 Bioanalyzer before sequencing. RNA was sequenced using the Illumina Genome Analyzer Ilx. The sequences obtained for the 24 libraries were converted to fastq format, and that is the format I used in my analysis.

The proposed MT sequence I generated was based on data from previous researchers in the laboratory (Thon et al., 2000; Thon et al., 2002), and the sequence of the plasmid used in the mutation experiments, pCB1636. I mapped the RNAseq reads from all the libraries to the proposed WT and MT sequences using BLAST (Basic Local Alignment Search Tool). The mapping of the sequences that matched was done using Gnumap with a 75% alignment score and the figures showing the hits were created using Integrated Genome Browser (IGB). Other figures were manually created using a vector graphics editor Inkscape.
AI.3 Results

Each of the five treatments that were sent for sequencing (WTAP, WTBT, WTNT, MTAP, MTBT) had three biological replicates, each in a separate, identical, labeled microfuge tube. Two technical replicates (lanes) were done for the WTAP, MTAP, and MTBT stages, which had very low fungal:plant ratios. WTBT and WTNT had only one technical replicate (lane). During my mapping, I discovered that one of the MTAP biological replicates contained no reads that matched the *Hyg* gene or the plasmid sequences, while one WTAP biological replicate did match those areas.

On Figure AI.2, the “X” marks where the plasmid had matches when using BLAST. The grey areas indicate the lanes that appear to be swapped. In biological replicate 1 for WTAP, in both technical replicates, I found reads that matched the plasmid sequence. The non-transformed WT strain should not contain any plasmid or *Hyg* sequences. PCR using primers against the *Hyg* gene confirmed that the WT does not contain it. Neither of the other two biological replicates of WTAP contains plasmid or *Hyg* sequences. Furthermore, although two biological replicates of MTAP did contain *Hyg* and plasmid sequences, biological replicate 3 had no matches to these sequences.

I have concluded that these two samples were probably switched, most likely due to a mixup with the tubes at Texas. Additionally, biological replicate 3 of the MTBT treatment also did not contain plasmid or hygromycin reads. Although this could just be a function of the relatively low number of hits I found overall in these regions, it could also be due to a mistake made by me and Dr. Torres in identifying samples here. It is possible that we used sheaths inoculated with WT instead of MT.

To address this problem, both lanes of biological replicate 3 of MTAP, and of biological replicate 1 of WTAP, were removed from the dataset. Furthermore, both lanes of biological replicate 3 of MTBT were also removed. The data were re-analyzed for my study as described in Chapter 2.

The corrected reads for the *Cpr1* gene and the two flanking genes are shown in Figure AI.3 for both fungal strains. The genes are numbered, and introns are
represented by the areas with no color. Reads from the WT are blue and from the MT are red.

Using the hypothesized mutant sequence (Figure AI.4 - Mutant), I mapped MT reads to the plasmid. Figure AI.4 shows the reads mapped to my original hypothesized mutant sequence (I ended up modifying this, as described in Chapter 2, based on my results). Identifying this error was very important for my analysis and for our future work with the transcriptome. However, our analyses of the WT data before and after the error was corrected revealed only a few, very minor differences. This is because it appears that, transcriptionally, the WTAP and MTAP stages are, statistically, virtually identical. Thus we do not believe that our original publication is going to cause any issues for anyone who may have already used the data for their own work.
Figure AI. 1 Wild-type (WT) Cpr1 gene (green) is represented with flanking genes (pink). This map of the mutant sequence shows the EcoRI sites (red boxes) and the Hygromycin and plasmid sequence areas as I believed them to be at the time.
<table>
<thead>
<tr>
<th>Lane</th>
<th>Treatment</th>
<th>Rep 1</th>
<th>Rep 2</th>
<th>Rep 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>LANE 1</td>
<td>WT AP</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANE 2</td>
<td>WT AP</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANE 3</td>
<td>MT AP</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LANE 4</td>
<td>MT AP</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LANE 5</td>
<td>WT BT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANE 6</td>
<td>MT BT</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LANE 7</td>
<td>MT BT</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>LANE 8</td>
<td>WT NT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure AI. 2 Technical replicates (Lanes), treatments and biological replicates (Rep) from the transcriptome sequencing project
Figure AI. 3 Transcriptome reads showing matches to the Cpr1 gene and the two flanking genes in the WT (blue) and the MT (red) strains.
Figure A1. 4 Hypothesized mutant sequence with reads matching the plasmid regions.
Figure AI. 5 Map of pAN56-1-sGFP-HDEL plasmid. Figure made with Geneious (version 6.0) created by Biomatters. Available from http://www.geneious.com
Figure A1. 6 Correlation analysis of fold changes by RNA sequencing and qRT-PCR. Log2 fold changes by qRT-PCR are plotted on the x-axis and RNAseq are plotted on the y-axis.

\[ y = 1.036x - 0.2391 \]

\[ R^2 = 0.8604 \]
Figure Al. 7 Corrected map of pCB1636. Figure made with Geneious (version 6.0) created by Biomatters. Available from [http://www.geneious.com](http://www.geneious.com).
Optimization of the protocol for race tube growth assays

AIII.1 Background

For my experiments in Chapter 4, I wanted to compare the growth of the MT and WT strains during exposure to a wide range of concentrations and types of chemical inducers of stress. Growth assays commonly described in the literature include liquid cultures (Angelova et al. 2005; Pakula et al. 2003); radial growth measurements (D. Li et al. 1995); and spore germination assays (Angelova et al. 2005). Spore germination assays seemed to be too labor-intensive to be feasible for the very large number of treatments I wanted to do (Slawecki et al., 2002). Liquid cultures required me to use too much of the chemicals, some of which were quite expensive. Thus I settled on radial growth measurements for my experiments. Radial growth is commonly used for comparing different fungal species and/or strains on different substrates or treatments (P. J. G. M. de Wit et al. 2012; Reeslev and Kjoller, 1995), including comparisons of resistance to different chemicals (Zheng et al. 2012).

In my first experiments, I grew the cultures on Petri plates on which I had overlaid a thin layer of agar containing the stress-inducing chemical. I used the agar overlays because it allowed me to use less of the chemicals. I took three measurements of mycelial growth for each plate at different locations, and then compared the average for the control with the averages for the different treatments. Unfortunately I found that the variation in my measurements with this technique was unacceptably large. I considered that the problem was probably the overlay: it was difficult to obtain an even layer of agar, and so some areas may have had more chemical than others.

Race tubes have long been used for studies of hyphal growth and circadian rhythms in *Neurospora crassa* (Ryan et al., 1943; Sargent, Briggs et al., 1966; Davis and Perkins, 2002). The races tubes are made of Pyrex glass. They are
not readily available, and they are also rather difficult to clean out after use. White and Woodward (1995) proposed a different method of producing race tubes by using plastic pipettes, which allows easy filling with media as well as being disposable. I needed an assay that would be accurate and repeatable, so I decided to compare the performance of standard Petri dishes to the race tubes for radial growth assays.

AIII.2 Material and Methods

AIII.2.1 Fungal strains

I used M1.001 Colletotrichum graminicola wild-type strain (L. J. Vaillancourt and Hanau 1991), as well as the 6-2 MT and MT-C complement strain derived from M1.001 (Thon et al. 2002), as described in Chapter 2.

AIII.2.2 Assays

Two assays were compared, one using Petri plates and the other using race tubes. Modified Fries Minimal (FM) agar media (Tuite 1969) was used for both assays. The medium consisted of 30 g of sucrose, 5 g ammonium tartrate, 1 g NH$_4$NO$_3$, 1 g KH$_2$PO$_4$, 0.48g MgSO$_4$.7H$_2$O, 1 g NaCl and 0.13 g CaCl$_2$.2H$_2$O. For the experiments to compare the performance of the Petri plates with the race tubes, I added tunicamycin, a chemical that causes secretion stress, to the media (as described in Chapter 4). Concentrations used were 0, 5, 10, 15 and 20 µg/ml. For the experiments to optimize the performance of the race tubes, un-amended FM was used, with no chemicals added.

For the Petri plates, 20 ml of FM was poured and solidified in the hood. A 5 ml overlay of molten agar containing tunicamycin was added and allowed to solidify. A small plug of mycelia was put in the middle of the plate, and it was incubated at 23°C. After 7 days I took three measurements of the colony radius on each Petri plate and calculated the average. For the race tube assays, I followed the protocol described by White and Woodward (1995). Twenty-five ml sterile disposable polystyrene pipettes (USA Scientific) were completely filled with molten FM medium (with or without tunicamycin) until it reached the maximum volume (approximately 36 ml). Then, the media was released until only 10 ml remained and the pipette was braced in a horizontal position until
the media solidified. The tip was removed, and the pipette was cut in half, by using a heated scalpel blade.

I did several experiments to optimize the performance of the race tubes. Different materials were tested for closing the tips of the race tubes. I used aluminum foil, aluminum foil held with a rubber band, Parafilm M (Pechiney Plastic Packaging Company) and Breathe-Easy sealing membrane (Sigma-Aldrich). For the race tubes, I also tested other parameters including location (open shelves, or contained inside a box with wet paper towel inside to maintain humidity); light and dark (transparent box, or in one covered with aluminum foil); volume of media (10 or 13 ml). I also evaluated the effect of different shelves inside the 23°C room, and also the region in the Petri dish from which the agar plug was removed (the edge of the colony or the middle).

To inoculate the race tubes, I used a 4mm cork borer to remove mycelial plugs from a plate colony, and with a scalpel I removed as much of the agar as I could. The mycelia plug was then inserted into one of the open end of the tube, approximately 2 mm from the end. The tubes were then sealed in some manner (see below) and incubated for 7 days. Linear growth was measured in just one direction, lengthwise along the tube.

The petri plates and race tubes were maintained under continuous light at 23°C for 14 days, when measurements of the linear mycelial growth were made. As the MT strain grows slightly slower than WT and C strains (Figure 1.3A), the difference in growth on the different treatments was always compared to the control of each strain, therefore creating a percentage of growth rate.

**All.3 Results**

In the comparison Petri plate assay, the main problem I had was the culture radius was not homogenous (Figure All.1A). That caused a lot of variation between the three different measurements (Figure All.1B), and produced a high standard deviation (Figure All.1C). This variation, together with the difficulty in producing an even overlay of the chemical medium, made this method problematic.

The race tubes, on the other hand, using the same chemical concentrations, showed less variation between the repetitions (Figure All.2A). Additionally the
race tubes allowed for a much longer incubation time, 14 days when compared to 7 days using the Petri plates. I noted that the fungus was more sensitive to the chemicals in the race tubes, compared with the Petri plates, for all treatments tested (DTT, tunicamycin, Paraquat, menadione).

Once I decided to use the race tubes for the growth experiments, I addressed several parameters in an effort to further diminish variation.

In two independent tests, any of several different ways to close the open sides of the tubes produced similar results. Aluminum foil, aluminum foil with rubber band, Parafilm M, and Breath-right were not statistically different by the Waller-Duncan test. Parafilm was chosen because it was the easiest to seal, avoiding the contamination that occasionally happened when using aluminum foil.

In regard to where the tubes were placed, either on open shelves in the 23°C room, or inside an open box or a closed one in the same room, there was also no difference between the treatments using the Waller-Duncan test. For convenience, I decided to put the tubes on the open shelves in the 23°C chamber. There was no difference in fungal growth or percentage media weight loss whether I used 10 ml or 13 ml, so I choose to use 10 ml, reducing the amount of chemicals needed. Only one of the factors I tested, the area of the Petri dish from which the mycelial plug was removed, had a significant effect on mycelial growth. Based on those experiments, I consistently used mycelial plugs removed from the edges of the colonies.

Based on all the optimization experiments I came up with the following for the race tubes: i) 10 mls of media in each pipette, producing 2 race growth tubes per pipette; ii) removing the mycelial plug with a cork borer from the edge of the colony to guarantee uniformity; iii) close both open sides of the tubes with parafilm; and iv) setting the tubes always on the same shelf of the 23°C growth chamber. Those parameters were used on the stress growth assays described in Chapter 3.
Figure AIII.1 Variability in radial growth during chemical stress. A) Mutant strain with the tunicamycin treatment (Control, 5, 10, 15 and 20 ug/ml), B) Drawing showing how the 3 measurements were taken, C) Measurement means in mm and standard deviation of *in vitro* growth for each strain in different tunicamycin concentrations.
Figure AIII.2 Race tubes growth. A) Linear growth means in cm and standard deviation of in vitro growth results for each strain in different tunicamycin concentrations using race tubes, B) Race tube with M1001 in control media.
Figure AIII.3 Alignment between *S. cerevisiae* Kar2/BiP (YJL034W) and *C. graminicola* homolog GLRG_10629.
Figure AIII.4 Primers and sequencing results of the promoter region used to produce a reporter construct for BiP using Gateway. Primer regions are underlined and the start codon of the gene is in bold.


Delivery, Functional Redundancy and Durable Disease Resistance.”

*Current Opinion in Plant Biology* 11: 373–79.


Chakrabarty, Romit, Rituparna Banerjee, Sang-Min Chung, Mark Farman, Vitaly Citovsky, Saskia a Hogenhout, Tzvi Tzfira, and Michael Goodin. 2007. “PSITE Vectors for Stable Integration or Transient Expression of Autofluorescent Protein Fusions in Plants: Probing Nicotiana
Chasman, Deborah, Yi-hsuan Ho, David B Berry, Corey M Nemec, Matthew E Macgilvray, James Hose, Anna E Merrill, et al. 2014. “Pathway Connectivity and Signaling Coordination in the Yeast Stress-Activated Signaling Network,” 1–23.


Gething, M J. 1999. “Role and Regulation of the ER Chaperone BiP.” 


Paczkowski, Jon E., and J. Christopher Fromme. 2014. “Structural Basis for Membrane Binding and Remodeling by the Exomer Secretory Vesicle


Following Host Jumps in the Irish Potato Famine Pathogen Lineage.”
doi:10.1126/science.1193070.


Rep, M, V Reiser, U Gartner, J M Thevelein, S Hohmann, G Ammerer, and H Ruis. 1999. “Osmotic Stress-Induced Gene Expression in


Schröder, Martin, Robert Clark, and Randal J. Kaufman. 2003. “IRE1- and HAC1-Independent Transcriptional Regulation in the Unfolded Protein

229


Thon, M R, E M Nuckles, and L Vaillancourt. 2000. “Restriction Enzyme-Mediated Integration Used to Produce Pathogenicity Mutants of


Yoshino, Kae, Hiroki Irieda, Fumie Sugimoto, Hirofumi Yoshioka, Tetsuro Okuno, and Yoshitaka Takano. 2012. “Cell Death of Nicotiana Benthamiana Is Induced by Secreted Protein NIS1 of Colletotrichum Orbiculare and Is Suppressed by a Homologue of CgDN3.” *Molecular


PLACE OF BIRTH
Lavras, Brazil

EDUCATION
B.S. Agronomy Engineer. 2006. Universidade Federal de Uberlandia. Brazil

PROFESSIONAL EXPERIENCE
Aug/2009-Apr2015. Graduate Research Assistant, University of Kentucky. USA.
2007. Visiting Scholar, University of Kentucky. USA.
2006. Field Assistant, Monsanto do Brasil Ltda. Brazil.

RESEARCH PUBLICATIONS

BUIATE, E.A.S.; Souza, E.A.; Vaillancourt, L.; Resende, I.; Klink, U.P.

BUIATE, E.A.S.; Brito, C.H.; Batistella, R.A., Brandão, A.M. Reação de híbridos de milho e levantamento dos principais fungos associados ao complexo de

Book Chapters
Crouch, J., O’Connell, R., Gan, P., BUIATE, E., Torres, M., Beirn, L., Shirasu, K.;

BUIATE, E.A.S.; Barcelos, Q.L.; Pinto, J.M.; Souza, E.A.; Venard, C.;

PUBLISHED CONFERENCE PROCEEDINGS
Conference Proceedings (8 total)
Abstracts (19 total)
Meeting Presentation with no published proceedings (1 total)

PROFESSIONAL SERVICE
Student representative on the Department of Plant Pathology Academic Program Committee, Jun/2013 – Apr/2015.

PROFESSIONAL ACTIVITIES
Member of GEN (Genetic Studies Group) at UFLA. Brazil. 2007-2009.
Member of APPS (Association of Plant Pathology Scholars) at UK. USA. 2010-2015.
Member of professional societies: The American Phytopathological Society and Mycological Society of America.

Ester Alvarenga Santos Buiate