Genomic Analysis, Population Quantification and Diversity
Characterization of Cryptococcus flavescens

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy
in the Graduate School of The Ohio State University

By
Xiaoqing Rong, M.S.
Graduate Program in Plant Pathology

The Ohio State University
2013

Dissertation Committee:
Dr. Brian B. McSpadden Gardener, Advisor
Dr. Thomas K. Mitchell
Dr. Pierce A. Paul
Dr. Christopher G. Taylor
Abstract

*Cryptococcus flavescens* strain OH182.9_3C (3C) has been shown to have biocontrol efficacy against Fusarium head blight (FHB) of wheat, and, 3C is currently under development as a biopesticide. The research described in this dissertation advanced that commercial development of this agent by providing new information about the biology, genetics, diversity, and ecology of this species for the control of Fusarium Head Blight. Such data may be used by the US EPA and the company developing the strain for the regulatory review of 3C as a new biopesticide ingredient.

A draft genome sequence of 3C was obtained through high-throughput sequencing and short read assembly. *Ab initio* gene prediction and sequence similarity search were conducted to annotate the genome. Nine conserved regions were identified from the assembled genome sequence to develop molecular markers for population quantification and diversity characterization. Putative non-ribosomal peptide synthetase and glucanase genes were also found in the genome and may contribute to the biocontrol activity previously noted for the strain.

A quantitative PCR (qPCR) assay was developed to quantify the abundance and population dynamics of 3C-like *C. flavescens* in wheat production systems. Following application of 3C to flowering wheat spikes, the population of 3C-like *C. flavescens* in the inoculated areas remained relatively stable for the remainder of the growing season. The populations of 3C-like *C. flavescens* seem to have spread from inoculated areas to non-inoculated areas and increased over time. Populations decreased on threshed grains during storage, but persisted in the postharvest field during the winter and early spring following the growing season.

Detailed study on diversity within *C. flavescens* species has not been reported before. Multilocus sequence typing was performed on strain 3C and 12 other *C. flavescens*
strains using six apparently unlinked loci: rDNA-ITS, β-tubulin, chitin synthase 1, elongation factor 1, heat shock protein 70 kDa and a mostly anonymous region. The rDNA-ITS sequences were identical across all the 13 strains. The other five loci consistently revealed two genotypes, A (6 strains, including 3C and type strain NRRL Y-1401T) and B (7 strains), with high bootstrap support. This genotypic split was congruent to the mating type previously performed on some of the tested strains. Phenotypic variation noted using BIOLOG assays supported the grouping by genotype and revealed the slow assimilation of several carbon sources by two isolates of genotype A. In greenhouse bioassays, the strains of both genotypes showed biocontrol efficacy against Fusarium head blight, but genotype A as a whole showed slightly greater biocontrol efficacy than genotype B strains. This study provided insight into the phylogenetic relationships within \textit{C. flavescens} species and resources for developing genetic and phenotypic markers to differentiate subspecies of \textit{C. flavescens}.

This research paves the way for additional studies of \textit{C. flavescens} and their use as biopesticidal agents. Genome-wide discovery of genes involved in biocontrol of FHB by 3C may help reveal the biocontrol mechanisms expressed by \textit{C. flavescens}. The developed qPCR assay and MLST can be used to further characterize the abundance and diversity of natural and inoculated populations of \textit{C. flavescens}. Collectively, such studies will further facilitate the development and registration of 3C and possibly other \textit{C. flavescens} strains as biopesticides for use against FHB.
Acknowledgements

I would like to acknowledge my Ph.D. advisor Dr. Brian McSpadden Gardener for the mentoring and support during my graduate career. He gave me great amounts of advice that was very helpful for my career development and personal growth. I also thank him for giving me opportunities for working on the genomic and molecular sides of microbiology which I am most interested in.

I thank my Student Advisory Committee members for advising me throughout the Ph.D. program. I thank Dr. Paul for his advice and help on the projects we collaborated on, as well as his thorough comments on my thesis. I thank Dr. Mitchell and Dr. Taylor for all their advice and suggestions on my projects.

I thank Dr. Paul and Dr. David Schisler from USDA-ARS for the collaboration on my dissertation projects, especially their kind help and readily offers of collaboration efforts. I thank Dr. Michael Boehm for his efforts in discovering and initially characterizing the biocontrol activities of *Cryptococcus flavescens* OH182.9, work that laid the foundation for my thesis research. I also thank all the past and current members in McSpadden Gardener lab and collaboration labs who have assisted with my projects.
Vita

2009.......................................................... B.S. Honors Program of Life Sciences, China Agricultural University

2012.......................................................... M.S. Plant Pathology, The Ohio State University

2009 to present ........................................... Graduate Research Associate, Department of Plant Pathology, The Ohio State University

Publications


- Rong, X.; Paul, P.A.; Schisler, D.A.; McSpadden Gardener, B.B. 2013. A quantitative PCR assay useful for monitoring the field dispersal and persistence of Cryptococcus flavescens, a biocontrol yeast applied to wheat. Submitted for publication.


**Fields of Study**

**Major Field:** Plant Pathology
Table of contents

Abstract......................................................................................................................................................ii
Acknowledgements.......................................................................................................................................iv
Vita................................................................................................................................................................v
Table of contents..........................................................................................................................................vii
List of Tables ................................................................................................................................................xi
List of Figures ..............................................................................................................................................xii
Chapter 1: Literature review – Fusarium head blight and Cryptococcus flavescens
OH182.9_3C, a biological control agent against this disease .........................................................1
   Fusarium head blight .................................................................................................................................1
      Overview of Fusarium head blight ........................................................................................................1
      Biology of pathogen infection and mycotoxin production .................................................................2
      Economic loss due to Fusarium head blight .......................................................................................7
   Need for biological control of FHB .........................................................................................................7
      Inadequacy of cultural, chemical, and host resistance in controlling FHB ......................................8
      Biocontrol as a promising approach to FHB control ........................................................................9
   C. flavescens OH182.9 as a biocontrol agent against Fusarium head blight (FHB) of wheat ...............9
      Discovery, isolation and selection of a fungicide-resistant variant ...................................................9
      Reduction of FHB damages by OH182.9 and OH182.9_3C ...............................................................11
   Taxonomy, physiology and sexual reproduction of C. flavescens species ...........................................12
      Taxonomic placement ...........................................................................................................................13
      Physiology ............................................................................................................................................14
      Sexual reproduction ..............................................................................................................................14
   Need to understand more about the biology, ecology and diversity of C. flavescens .........................15
   Tables and Figures ..................................................................................................................................16
Chapter 2: Genomic Analysis of C. flavescens OH 182.9_3C................................................................20
Abstract.................................................................................................................................20
Introduction.............................................................................................................................20
  Need for genetic markers to detect and characterize 3C and other C. flavescens strains .................................................................20
  Types of biocontrol mechanism in general and yeast biocontrol agents for plant diseases .................................................................20
  Genomic analysis on C. flavescens OH182.9_3C (3C) can help to decipher its biocontrol mechanisms ..............................................23
  Identification of genetic markers to detect and characterize 3C and other C. flavescens strains .............................................................24
  Genome sequencing and assembly of 3C .................................................................................24
  Identification of conserved regions that can be used as genetic markers ..........25
  Primer design for quantitative PCR to detect 3C in the field .............................................26
  Primer design for multilocus sequence typing to characterize C. flavescens strains ........................................................................27
  Genome-wide annotation for 3C through JGI, Augustus and blastx ........................................28
Tables and figures ...................................................................................................................32
Chapter 3: Monitoring field dispersal and persistence of C. flavescens OH182.9_3C using quantitative PCR ..................................................39
Abstract..................................................................................................................................39
Introduction.............................................................................................................................40
Materials and methods ............................................................................................................42
  Strains..................................................................................................................................42
  Field establishment and 3C application ..............................................................................43
  Inoculum production ..........................................................................................................43
  Sample collection ..............................................................................................................44
  Sample handling, DNA extraction and template preparation ............................................45
  Quantitative PCR (qPCR) ..................................................................................................46
  Data processing and analysis ..............................................................................................46
Results ..................................................................................................................................48
  Sensitivity and specificity of the assay ..............................................................................48
  Dynamics of 3C-like C. flavescens population in areas inoculated with 3C ..........................................................49
  Dispersal of 3C to the non-inoculated area of field ..............................................................50
  The effect of sample collection activity on 3C dispersal .....................................................50
The effect of prevailing wind direction on 3C dispersal.................................51
Overwintering persistence of 3C-like C. flavescens ..................................52
Population levels of 3C-like C. flavescens around harvest and during storage
......................................................................................................................52
Discussion ........................................................................................................53
Effectiveness of the qPCR assay .................................................................53
Population dynamics of 3C-like C. flavescens in the field ..........................55
Possible application of the qPCR assay .......................................................56
Tables and Figures ..........................................................................................57
Chapter 4: Multilocus sequence typing and phenotyping of 15 C. flavescens strains......67
Abstract..........................................................................................................67
Introduction .....................................................................................................68
Materials and Methods ................................................................................69
Strains ............................................................................................................70
Identification of MLST loci and choice of primers ......................................70
DNA isolation, PCR and sequencing ..........................................................70
Phylogenetic analysis ....................................................................................71
Biochemical assay .......................................................................................71
Biocontrol assays ..........................................................................................72
Results ............................................................................................................74
Evidence for two distinct genotypes of C. flavescens ...............................74
Taxometric analysis on biochemical properties of C. flavescens strains shown in Biolog assay .................................................................75
Biocontrol efficacy of the two subspecies of C. flavescens suppressing
Fusarium Head Blight (FHB) ........................................................................76
Discussion ....................................................................................................77
Tables and Figures ..........................................................................................77
Chapter 5: Future research ...........................................................................80
Genome-wide discovery of genes involved in biocontrol of 3C against FHB ....95
Transcriptomic analysis of interactions among 3C, Fusarium graminearum
(Fg) and wheat ...............................................................................................96
Gene knockout analyses on the 3C genes potentially involved in biocontrol.97
Characterizing the diversity and abundance of natural C. flavescens populations in
wheat and barley ..........................................................................................99
Validity of using developed qPCR and MLST primers for future studies ..... 99
Characterizing the diversity and abundance of natural *C. flavescens*
populations using the developed qPCR and MLST primers .................. 101
Marker-assisted bioprospecting of *C. flavescens* isolates that are potential
biocontrol agents .............................................................................. 102
Marker-assisted isolation of *C. flavescens* isolates from wheat and barley
followed by biocontrol efficacy test .................................................. 102
Develop genotypic markers more indicative of biocontrol efficacy of *C.
flavescens* isolates ........................................................................... 104
Develop *C. flavescens*-selective media for bio-prospecting based on patterns
of carbon utilization and chemical resistance shown in Biolog assay ........ 104
Determining environmental fate of 3C-type biocontrol *C. flavescens* by monitoring
their dispersal and survival in production system of wheat and barley .......... 105
Breeding of *C. flavescens* for higher biocontrol efficacy ........................ 107
Identify mating loci and develop sequence typing methods to determine
mating types ..................................................................................... 107
“Yeast breeding” – Does hybrid progeny provide better biocontrol against
FHB? .................................................................................................. 108
Conclusions .......................................................................................... 109
References .......................................................................................... 110
Appendix A: computer scripts used in genomic analysis ......................... 123
Python script for converting multiple lines of sequences between two headers into
one line ............................................................................................... 123
Python script for splitting scaffolds at > 10 consecutive ‘N’s ....................... 124
Python script for making .agp file for your assembly .................................. 127
Sample .agp files for scaffolds (assembly with > 10 consecutive ‘N’) .............. 129
Sample .agp files for contigs (assembly without > 10 consecutive ‘N’) ........... 129
Customization of existing Python scripts for converting BLAST .xml into tab-
separated values (also known as .tsv, .tab or .txt) .................................. 130
Customization of ‘blast_xml.py’ ............................................................ 131
Customization of ‘blast_tab.py’ ............................................................. 132
List of Tables

Table 1. Summary of greenhouse data on biocontrol efficacy of OH182.9 against Fusarium head blight\textsuperscript{a} .................................................................................................................................................. 17
Table 2. Summary of field data on biocontrol efficacy of OH182.9 against Fusarium head blight\textsuperscript{a} .................................................................................................................................................. 18
Table 3. The conserved regions of 3C genome used for marker development ................................................................. 33
Table 4. The primers designed for quantitative PCR to detect 3C in field ......................................................................... 34
Table 5. The primers used for multilocus sequence typing to characterize \textit{C. flavescens} strains, all designed in this study except ITS5 and ITS4 (White 1990) ................................................................................................................................. 35
Table 6. Prediction of putative non-ribosomal peptide synthetase genes by Augustus, JGI and Blastx. ................................................................. 36
Table 7. Prediction of putative glucanase genes by Augustus, JGI and Blastx. ......................................................... 37
Table 8. Sensitivity and specificity of the qPCR assay targeting the putative \textit{hsp70} gene. ................................................................. 61
Table 9. \textit{C. flavescens} strains used in this study ................................................................................................................................. 81
Table 10. Primers and target loci for the MLST analysis on \textit{C. flavescens} strains ................................................................. 82
Table 11. DNA sequence identities among \textit{C. flavescens} strains ......................................................................................... 83
Table 12. Biochemical characteristics of \textit{C. flavescens} strains shown by OD595 of Biolog assay (see next page) ................................................................................................................................. 84
List of Figures

Figure 1. Venn diagram showing the numbers of *C. flavescens* 3C scaffolds where features were predicted by Augustus and/or JGI and/or Blastx hits were found. 38
Figure 2. Field map and major sampling scheme for the detection of *C. flavescens* strain 3C in Field One. 58
Figure 3. Sequence alignment of the qPCR amplicons from 3C and other six *C. flavescens* strains. 60
Figure 4. Population dynamics of 3C-like *C. flavescens* on inoculated wheat heads. 62
Figure 5. Changing populations of 3C-like *C. flavescens* in a wheat field following inoculation in part of the field. 63
Figure 6. Overwintering persistence of 3C-like *C. flavescens* in inoculated area. 65
Figure 7. 3C-like *C. flavescens* populations on wheat heads closely before harvest and postharvest threshed grains. 66
Figure 8. Genetic similarity among *C. flavescens* strains shown by the concatenated DNA sequences of β-tubulin, Chitin Synthase 1, Elongation Factor 1 and Heat Shock Prote in 70 kDa. 86
Figure 9. *C. flavescens* strains consistently split into two groups shown in individual DNA trees. (See next page) 87
Figure 10. Cluster analysis on the carbon source utilization and chemical sensitivity patterns of *C. flavescens* strains at two time points (A and B) in Experiment One. Type strain Y-1401T was not included. (See next page) 90
Figure 12. Cluster analysis on the carbon source utilization and chemical sensitivity patterns of type strain Y-1401T and other selected *C. flavescens* strains at two time points (left and right) in Experiment Two. 93
Figure 13. Biocontrol efficacy of two distinct genotypes of *C. flavescens*. 94
Chapter 1: Literature review – Fusarium head blight and
Cryptococcus flavescens OH182.9_3C, a biological control agent
against this disease

**Fusarium head blight**

**Overview of Fusarium head blight**

Fusarium head blight (FHB) is a devastating plant disease in humid and semi-humid regions throughout the world. FHB is also known as Fusarium head scab or Fusarium ear blight. The hosts of this disease include several economically important crops such as hexaploid wheat (*Triticum aestivum*), durum wheat (*Triticum durum*), barley (*Hordeum vulgare*) and oat (*Avena sativa*). Several species of *Fusarium*, including *Fusarium avenaceum*, *Fusarium culmorum*, *Fusarium graminearum*, and *Fusarium poae* cause symptoms of head blight (Parry, Jenkinson, & McLeod, 1995; Schmale III & Bergstrom, 2003; Wong, Tekauz, Leisle, Abramson, & McKenzie, 1992). However, the predominant pathogen of this disease in North America is *Fusarium graminearum sensu stricto* Schwabe (anamorph), of which the teleomorph is *Gibberella zeae* (Schwein.) Petch (Aoki & O'Donnell, 1999; O'Donnell, Ward, Geiser, Corby Kistler, & Aoki, 2004; O'Donnell, Kistler, Tacke, & Casper, 2000). This pathogen displays phenotypic and phylogenetic diversities across different geographic locations in North America (Ward et al., 2008), Australia (Akinsanmi, Backhouse, Simpfendorfer, & Chakraborty, 2006) and South Korea (Lee et al., 2009).

Infection of wheat heads by *F. graminearum* is initiated by ascospores or macroconidia disseminated to the wheat heads by wind, rain or insects (Sutton, 1982).
After landing, spores germinate and develop hyphae growing towards susceptible tissues such as openings (e.g. stomata) or thin-walled surfaces (e.g. anthers and underlying parenchyma) (Bushnell, Hazen, & Pritsch, 2003; Goswami & Kistler, 2004). Wheat heads are susceptible to infection from anthesis to the soft dough stage of kernel development (McMullen, Jones, & Gallenberg, 1997; Sutton, 1982). Favorable conditions for infection and disease development include frequent rainfall, high humidity and heavy dew formation during susceptible stages. Typical disease symptoms are bleaching of glume tissue and shriveled “tombstone” kernels. The pathogen produces mycotoxins such as deoxynivalenol and zearelanone. Deoxynivalenol, which causes vomiting and feed refusal by livestock, is also known as vomitoxin. Economic loss caused by this disease lies in decreased yield, reduction of seed quality and lower market grade of the grain due to Fusarium-damaged kernels and contamination with mycotoxins.

**Biology of pathogen infection and mycotoxin production**

Understanding of the infection and mycotoxin production processes of the pathogen *F. graminearum* is necessary for management of FHB in terms of discovering strategies to disrupt these processes.

**Inoculum**

The inocula of *F. graminearum* are ascospores and macroconidia. Ascospores are produced by sexual fruiting bodies called perithecia following mating of haploid parents and then meiosis in asci. They can be dispersed from perithecia, which form on overwintering residues, to host plants by wind, rain, insects or other disseminating agents (Sutton, 1982). It has been shown that ascospores were able to germinate in 3 to 6 hours in water that was not amended with any nutrients and that germination tubes could elongate without exogenous nutrients shortly after spore germination (Bushnell et al., 2003). Macroconidia are produced as asexual reproduction structures on vegetative mycelium and dispersed by similar means to ascospores. Germination of macroconidia was observed within 6 to 12 hours after inoculation at the abaxial side of glume on wheat.
heads and hyphae developed within 12 to 24 hours after inoculation (Pritsch, Muehlbauer, Bushnell, Somers, & Vance, 2000).

Penetration of host tissue

Hyphae grow rapidly following germination and usually form mycelium network on the surfaces of wheat glume and florets, and this was especially evident when large numbers of spores were used as inoculum (Pritsch et al., 2000). Under highly moist conditions such as in greenhouse or moist chamber, mycelium could even develop to such thickness that it was visible by unaided eye (Bushnell et al., 2003). In contrast, relatively scarce mycelia are usually observed at lower moisture such as in field. Pathogen hyphae are usually not able to directly penetrate the epidermis cells of wheat glume probably because of the thick and crenulated cell walls that contain high amounts of silicon deposition (Hodson & Sangster, 1988; Hodson & Sangster, 1989; Pritsch et al., 2000). Instead, it is widely believed that hyphae extend along the surface of glume in search of stomata openings and thinner-walled tissues such as anthers, and crevices between palea and lemma, which have weaker barrier than thick-walled epidermis. In addition, loss of wax and possibly cuticle from the surfaces of lemma and palea was observed around hyphae under scanning electron microscope (Bushnell et al., 2003; Kang & Buchenauer, 2000b). This suggests that the pathogen may be able to utilize the wax tissue as nutrient sources and/or dissolve it to increase the permeability of the cuticle to more readily absorb nutrients from wheat cells underneath. Subcuticular growth on wheat glume has been observed for F. graminearum and another pathogen of Fusarium head blight, F. culmorum (Kang & Buchenauer, 2000b; Pritsch et al., 2000). F. graminearum was able to produce a network of abundant subcuticular hyphae on the abaxial surface of wheat glumes. The hyphae tended to extend in a more rapid fashion along the rows of stomata and showed a highly branched coralloid pattern (Pritsch et al., 2000).

How the pathogen penetrates the surface of wheat tissue and gains access to the interior has been addressed by many studies. Wound is a possible avenue for entry. Damage due to stalk rot of maize, also caused by F. graminearum, were increased by artificial wounding, wind gusts, or insect feeding (Christensen & Wilcoxson, 1966).
However, pin prick wounds made on the exterior lemma surface of wheat spikes did not facilitate the growth of pathogen into the interior tissue (Bushnell et al., 2003). Entry through stomata has been demonstrated using microscopic methods (Kang & Buchenauer, 2000b; Pritsch et al., 2000). The thin-walled chlorenchyma tissue located beneath stomata probably can be readily penetrated by the pathogen and provide nutrients for further development of hyphae through active photosynthesis (Bushnell et al., 2003).

Anthers were shown to be important avenue for *F. graminearum* infection. Partially exserted anthers were observed to be heavily colonized by the pathogen (Pugh, Dickson, & Johann, 1933). Complete removal of anthers before inoculation with the pathogen resulted in less frequent infection (Bushnell et al., 2003). Two ammonium compounds choline and betaine were identified in anther extracts and demonstrated promoting effects on linear growth of *F. graminearum in vitro*, therefore they were considered as enhancing agents of pathogen virulence (Strange, Majer, & Smith, 1974; Strange, Deramo, & Smith, 1978; Strange & Smith, 1978a; Strange & Smith, 1978b). Pollen was also believed to facilitate disease development. Abundant *F. graminearum* growth was observed on the surfaces of pollen (Ribichich, Lopez, & Vegetti, 2000) and pollen dialysates were shown to increase germination rates of pathogen spores and elongation of germination tubes (Bushnell et al., 2003).

**Degradation of host cell wall**

*F. graminearum* and other *Fusarium* species are known to produce enzymes that degrade cell wall components of wheat hosts. At least 8 isolates of *F. graminearum* were able to produce cellulase and polygalacturonase on cultivar Fortuna of wheat (Bushnell et al., 2003). *F. culmorum* infection resulted in reduction of cellulose, xylan and pectin in host tissue near intercellular hyphae and penetration pegs, indicating the degradation activity of the pathogen (Kang & Buchenauer, 2000a).

**Short biotrophic stage before necrosis**

*F. graminearum* does not cause cell death of host immediately after colonization. The pathogen has been isolated from asymptomatic tissues of infected wheat plants such as
rachis and kernel (Bushnell et al., 2003), which indicates that it is able to survive on host without causing disease. Furthermore, glume tissue inoculated with the pathogen remained healthy-looking without chlorosis or other visible damages 1 to 3 days after inoculation, even when intercellular growth of hyphae was observed in mesophyll (Bushnell et al., 2003; Pritsch et al., 2000). Mesophyll started to lose chlorophyll approximately 1 mm behind the advancing front of intercellular hyphae and turned chlorotic to brown by 3 to 4 days after inoculation, when intracellular hyphae were observed in cell lumens of epidermis and mesophyll (Bushnell et al., 2003). Infection by GFP-labeled F. graminearum showed that cell death is induced as soon as hyphae invaded the cytosol of epicarp cells, therefore an intracellular biotrophic stage was not present (Jansen et al., 2005). This indicates that F. graminearum is biotrophic intercellularly at the early stage of infection before it becomes necrotrophic intracellularly in invaded cells.

*Mycotoxin production and its role in pathogenesis*

*F. graminearum* is known to produce mycotoxins such as deoxynivalenol (DON), 3-Acetyldeoxynivalenol (3-A DON), 15-Acetyldeoxynivalenol (15-A DON), nivalenol and zearalenone (Bily et al., 2004; Mirocha, Yu, Evans, Koloczowski, & Dill-Macky, 1997). DON family and Nivalenol are the most abundant trichothecene toxins found in wheat grains infected by *Fusarium* species. Trichothecene compounds are characterized by chemical structures of a tricyclic ring, a double bond at C-9,10 and an epoxide at C-13. About half of the genes responsible for the biosynthesis of trichothecenes are located in a 25-kb region “Tri cluster”, which include 14 genes *Tri1* to *Tri14* (Kimura, Anzai, & Yamaguchi, 2001). Different strains of *Fusarium* species showed different patterns of trichothecene toxin production and thus were classified into different chemotypes (Goswami & Kistler, 2004). The chemotypes are not well correlated with phylogenetic classifications (O'Donnell et al., 2000).

Trichothecene toxins have multiple effects on both plant and animal cells, including inhibition of protein synthesis, inhibition of DNA and RNA synthesis, alteration of membrane structure, and induction of programmed cell death (McCormick, 2003; Rocha,
As phytotoxins, trichothecenes can cause chlorosis, necrosis and wilting of plants, which contribute to the symptoms caused by *F. graminearum* such as bleaching and browning of wheat head tissues. Accumulation of deoxynivalenol increased as soon as pathogen started active development and infection in host tissue than after inoculation (Boddu, Cho, Kruger, & Muehlbauer, 2006). Reduced disease incidence and disease severity was observed in field-grown wheat infected with trichothecene non-producing mutant *Tri5* compared to wild type *F. graminearum* (Desjardins et al., 1996). The trichothecene non-producing mutant was able to invade the inoculated wheat florets but unable to spread to non-inoculated ones, suggesting that trichothecene toxins are necessary for spread of the pathogen beyond penetration sites (Eudes, Collin, Rious, & Comeau, 1997; Mirocha et al., 1997).

Comparison of GFP-labeled wild type and a trichodiene synthase disrupted mutant showed different progression patterns between the wild type and the mutant (Jansen et al., 2005). After germination from spores on wheat head tissues, hyphae need to penetrate multiple layers of the floret to reach the endosperm rich in starch and protein. Wild type and mutant showed no difference in ability to destroy the different layers and infect the endosperm. After destroying the floret, wild type progressed through rachis to the next apically located floret, while the mutant was contained in the original floret and heavy thickening of cell wall at rachis node was observed. This indicated that trichothecene toxins suppress cell wall fortification at rachis node as a host defense mechanism thus facilitate the spread of pathogen from infected floret to neighboring healthy florets (Jansen et al., 2005).

**Transcriptomic changes of *F. graminearum* during infection**

Transcriptomic analyses on *F. graminearum* has been conducted in time series during disease development for Fusarium head blight as well as crown rot of wheat, another disease caused by *F. graminearum* (Lysøe, Seong, & Kistler, 2011; Stephens, Gardiner, White, Munn, & Manners, 2008). Differentially expressed metabolism-related genes *in planta* versus on axenic media included the following functional categories: carbohydrate metabolism, degradation of polysaccharides and ester compounds, secondary
metabolism, nitrogen, sulfur and selenium metabolism. Differential expression was also detected in many other categories of genes such as cell cycle and DNA processing, protein synthesis and fate, cellular transport (e.g. allantoin and allantoate transport), detoxification, cell type differentiation, cell rescue and defense, and secreted virulence factors.

**Economic loss due to Fusarium head blight**

The damage caused by Fusarium head blight is multi-folded. Symptomatic wheat heads yield less grain, in addition, infected grains are shriveled and light-weighted (“tombstone” symptom). The grains are also contaminated with mycotoxins produced by the pathogen. Deoxynivalenol (DON), also known as vomitoxin, is one of the major mycotoxins found on *Fusarium*-infected wheat grains. DON-contaminated grains cause vomiting and feeding refusal by non-ruminant livestock. Therefore, infected wheat grains are of much lower market values and frequently experience substantial discount and even rejection in trading.

The economic loss due to Fusarium head blight can be tremendous. The 1993 epidemic in a tri-state area of Minnesota, North Dakota and South Dakota, one of the largest wheat and barley production region in US, the loss due to yield reduction and price discount was approximately $1 billion. This was considered as one of the greatest losses due to any plant disease in North America in a single year (McMullen et al., 1997). In the 1996 epidemic that affected 10 states in the northeast US, reduced yield and quality caused an estimated loss of $40 to 100 million for each state (McMullen et al., 1997). Losses due to epidemics in 9 northeast states between 1998 and 2000 were almost $2.7 billion, including direst loss in yield and price as well as economic impact (Nganje et al., 2001; Nganje & Johnson, 2003). And the loss of business activity would have supported up to 14 thousand full-time equivalent jobs in each year from 1998-2000 (Nganje et al., 2001).

**Need for biological control of FHB**
Inadequacy of cultural, chemical, and host resistance in controlling FHB

Currently, Fusarium head blight (FHB) is controlled mainly through chemical fungicides. Triazole-based fungicides have been shown to be most effective and consistent among various families of fungicides (Mesterházy, Bartók, & Lamper, 2003; Paul et al., 2008). Tiazole fungicides were able to reduce disease index by 19% to 63%, DON accumulation by 7% to 58% and yield loss by 21% to 60%, depending on wheat cultivar as well as year and location of field trials (Beyer, Klix, Klink, & Verreet, 2006; Mesterházy et al., 2003; Mesterházy, 2003; Paul et al., 2008). However, complete control of FHB has not been achieved and high variability in efficacy was often observed. Furthermore, triazole-based fungicides are listed as possible carcinogens (category C) by US Environmental Protection Agency (List of Chemicals Evaluated for Carcinogenic Potential, Office of Pesticide Programs, US Environmental Protection Agency), which means they have potential risks to human health.

Resistance to FHB is classified into two main types. Type I resistance reduces penetration and establishment of the pathogen, Type II resistance suppress progression of the pathogen from infected tissue to healthy tissue, and Type III resistance decreases the accumulation of DON. Various mechanisms of resistance to FHB have been identified in host. Salicylic acid is a key phytohormone that regulates basal resistance of wheat to FHB (Makandar et al., 2012). Production of antifungal proteins such as pathogenesis-related (PR) proteins is one of the basal defense mechanisms against general Fusarium infection with low specificity to pathogen isolates (Pritsch et al., 2001). Anther extrusion was also reported to be associated with resistance to FHB (Skinnes, Semagn, Tarkegne, Marøy, & Bjørnstad, 2010). In addition, multiple quantitative trait loci in wheat were related to FHB resistance (Anderson et al., 2001; Beyer et al., 2006). However, wheat breeding lines with high levels of resistance usually cannot compete with commercial cultivars in terms of agronomic performance such as yield and complete resistance has not yet been achieved (Beyer et al., 2006).

Most of FHB inocula are ascospores and macroconidia produced on crop residues, therefore reducing infected crop residues above or right below soil surface will reduce the
chances of disease occurrence in the following growing season. Conventional tillage to bury residues deep under soil surface and burning of residues was shown to significantly reduce disease levels (Beyer et al., 2006; Dill-Macky & Jones, 2000; Lori, Sisterna, Sarandón, Rizzo, & Chidichimo, 2009). However, conventional tillage results in such substantial loss of soil organic matters and water-stable aggregates that reduced or no tillage is often recommended for soil conservation (Kushwaha, Tripathi, & Singh, 2001; Leys, Govers, Gillijns, Berckmoes, & Takken, 2010). The inocula spread from nearby fields also contribute to the difficulty of controlling the disease by inoculum removal.

**Biocontrol as a promising approach to FHB control**

Biological control is an environment-friendly strategy for control of crop diseases and pests. It utilizes microorganisms and/or their metabolites to suppress plant pathogens and/or to promote plant growth. Multiple species of bacteria and fungi have shown biocontrol activities against FHB (Corio da Luz, Stockwell, & Bergstrom, 2003), however, none was able to completely control FHB whether applied alone or in combination with other disease management strategies. Therefore, more efforts towards discovery and development of novel biocontrol agents against FHB are demanded.

**C. flavescens** OH182.9 as a biocontrol agent against Fusarium head blight (FHB) of wheat

**Discovery, isolation and selection of a fungicide-resistant variant**

*Cryptococcus flavescens* OH182.9 (NRRL Y-30216) was isolated as an anther colonist of wheat (Khan, Schisler, Boehm, Slininger, & Bothast, 2001). More than 400 wheat anthers were collected from wheat plants across Illinois and Ohio in the spring of 1997, where and when FHB epidemics had recently occurred. Anthers were collected in vials containing 10% (v/v) glycerol and kept at 5°C in the field, then stored at -80°C prior to further processing. Vials were vortexed for 30 s to release microorganisms from anther
surfaces, and then suspensions were serially diluted in sterile, weak, pH 7.2 phosphate buffer. Dilutions were then plated onto a variety of solidified media with different selecting preferences. A total of 738 microbial isolates were purified, of which 54 tartaric acid-utilizing strains and 188 non-utilizing strains were subject to assay for biocontrol efficacy. Eight tartaric acid-utilizing strains and 18 non-utilizing strains prevented any visible FHB disease development in a two-head plant bioassay. Of the 26 effective strains, 4 tartaric acid-utilizing strains (all were yeasts, Cryptococcus or not determined) and 3 non-utilizing strains (all Bacillus) reduced FHB severity by 25% to 80%. OH182.9 was one of the four tartaric acid-utilizing yeasts.

The 7 antagonists were then assayed for biocontrol efficacy against different F. graminearum isolates and to evaluate the effects of concentration and timing of inoculum application (Khan et al., 2001). Depending on the F. graminearum isolate tested, OH182.9 reduced FHB disease severity by 6% to 57% and disease incidence by 2% to 24%, and increased 100 g kernel weight by 9% to 100%. Reduction of disease severity and incidence by OH182.9 was significant for F. graminearum isolate Z3639 but not for DAOM 180378 or Fg-9-96, while increase of kernel weight was significant for all the three isolates. OH182.9 was more successful in reducing FHB severity when applied as a mist than as a single point on wheat heads. Interactions between OH182.9 inoculum concentration, timing of OH182.9 application, and F. graminearum inoculation were observed for reduction of disease severity and increase of kernel weight. Among the conditions tested, better disease control was obtained when OH182.9 was applied 4 h prior to pathogen (Z3639) inoculation than when OH182.9 was applied immediately before, immediately after or 4 h after pathogen inoculation. Effect of OH182.9 inoculum concentration depended on whether disease severity or kernel weight was evaluated. OH182.9 did not demonstrate antagonistic activity against the pathogen in an in vitro co-culture assay on petri dish (Khan et al., 2001).

Since OH182.9 was not able to completely control FHB, its application in combination with other disease control methods such as fungicides is desired. To this end, selection for OH182.9 variants that can tolerate fungicides used for FHB control was performed through protracted exposure of OH182.9 to Proline®, a prothioconazole-based
fungicide against FHB (Schisler, Boehm, & Paul, 2009). One of the variant isolates OH182.9_3C (NRRL Y-50378) was shown to reduce disease severity by 34% in two field trials on susceptible winter wheat. For both trials, treating wheat heads during anthesis with Proline® followed by OH182.9_3C 24 h later was the most successful treatment in reducing FHB severity (control vs. treated average severity of 10.0% vs. 1.4%, respectively) and deoxynivalenol accumulation (10.0 vs. 2.3 ppm, respectively).

**Reduction of FHB damages by OH182.9 and OH182.9_3C**

OH182.9 has been shown to substantially reduce damage caused by FHB in both greenhouse and field experiments. *F. graminearum* isolate Z3639, a producer of 15-Acetyldeoxynivalenol (15-A DON) (Kang & Buchenauer, 2000b), was used in most of the trials. In greenhouse, OH182.9 was able to reduce disease severity and disease incidence by up to 57% and 24%, respectively, and to increase yield (measured by 100 kernel weight) by up to 100%, compared to negative control (buffer used to deliver OH182.9 inoculum) (Table 1) (Khan, Schisler, & Boehm, 1999a; Khan, Schisler, & Boehm, 1999b; Khan et al., 2001; Schisler, Khan, Boehm, & Slininger, 2002). In field experiments across different states including Ohio, Illinois and North Carolina, OH182.9 was able to reduce disease severity, disease incidence and deoxynivalenol accumulation by up to 70%, 60% and 42%, respectively, and to increase yield (measured by 100 kernel weight) by up to 8%, compared to negative control (buffer used to deliver OH182.9 inoculum) (Table 2) (Boehm, Khan, & Schisler, 1999; Khan, Schisler, Boehm, Lipps, & Slininger, 2004; Schisler, Boehm, & Lipps, 2000; Schisler et al., 2002). This indicates that OH182.9 is a promising biocontrol agent against FHB.

However, the efficacy of OH182.9 is greatly affected by multiple predictor variables due to its nature of being an organism able to interact with various factors. It showed different levels of control against different *F. graminearum* isolates in green house, e.g. reduction of disease severity and disease incidence by 56.7% vs. 5.6% and 24.2% vs. 5.1%, respectively, and increase of yield 100% vs. 9.4% for Z3639 vs. Fg-9-96, respectively (Table 1) (Khan et al., 1999b; Khan et al., 2001). Different wheat cultivars responded to OH182.9 very differently as well. Yield was increased by OH182.9 on some
cultivars but decreased on others in both greenhouse and field trials. Similar trends were observed for disease severity and disease incidence in field (Table 1, Table 2) (Khan et al., 2004; Schisler et al., 2000; Schisler et al., 2002). Depending on the application timing versus pathogen inoculum in the greenhouse, reduction of disease severity and increase of yield varied from 25.6% to 43.0% and 0 to 92.9% (Table 1) (Khan et al., 1999b; Khan et al., 2001). The effect of OH182.9 inoculum concentrations ranged from reduction to increase of disease damages in field as well (Table 2) (Khan et al., 2004). Carbon to nitrogen ration of the media used for OH182.9 inoculum production had relatively little effect on efficacy, but around a 2-fold difference in reduction of disease severity or disease incidence was occasionally observed in both greenhouse and field (Khan et al., 2004; Schisler et al., 2000; Schisler et al., 2002). This indicates that optimization of formulation and application practice is necessary to achieve maximum and consistent efficacy of OH182.9.

OH182.9_3C (3C), the prothioconazole tolerant variant of OH182.9, has been tested in combination with Prosaro®, a prothioconazole-based fungicide against FHB. Based on two years of field trials, 3C plus full strength Prosaro® performed better at reducing FHB severity and incidence as well as increase of yield than 3C plus 1/10 strength Prosaro®, 3C or Prosaro alone, or wild type OH182.9 (Schisler, Boehm, Paul, & Dunlap, 2009; Schisler, Paul, Boehm, Bradley, & Dunlap, 2010). This indicates that 3C can be used in addition to Prosaro® in normal wheat production to achieve better control of FHB. Culture-based quantification by colony forming units showed that 3C population was lower when applied in combination with full strength Prosaro® than when combined with 1/10 strength Prosaro® or 3C alone. This suggests that Prosaro® was limiting the growth of 3C population. Improving formulation of 3C as a biopesticide may increase the rates of 3C population growth/survival so that 3C can perform even better when used together with Prosaro®.

**Taxonomy, physiology and sexual reproduction of C. flavescens species**
Taxonomic placement

The species of *Cryptococcus flavescens* was first reported in 1922 as an isolate from the atmosphere in Japan (Saito, 1922). Back then it was considered as belonging to an Ascomycete genus *Torula* and named as *Torula flavescens* Saito. Later it was considered as a synonym to *Torulopsis flavescens* (Saito) Lodder, of which the genus *Torulopsis* is now *Candida*, also an Ascomycete yeast (Lodder, 1934). Yet later it was changed to the Basidiomycete genus *Cryptococcus* and re-named as *Cryptococcus laurentii* (Kufferath) C. E. Skinner var. *flavescens* (Saito) (Lodder & Kreger-van Rij, 1952).

It was not until recently that DNA sequence and protein evidence indicated that *C. laurentii* var. *flavescens* represents a separate species than *C. larentii* (Fonseca, Boekhout, & Fell, 2011; Takashima, Sugita, Shinoda, & Nakase, 2003). Phylogenetic analysis on sequences of 18S rDNA and the D1/D2 region of 26S rDNA showed that it was in the same clade as *C. aureus* (previously *Rhodotorula aurea*, a separate species than *C. laurentii*) while in a different clade than *C. laurentii* (Takashima et al., 2003). It was then reinstated as *Cryptococcus flavescens* (Saito) C. E. Skinner and but remained in the *Bulleromyces* clade of the Tremellales order (Takashima et al., 2003). The *C. flavescens* strain originally isolated in 1922 became the type strain CBS 942\(^T\) = ATCC10668\(^T\) = DBVPG 6007\(^T\) = MUCL 30414\(^T\) = NRRL Y-1401\(^T\). In terms of phylogenetic distance based on the D1/D2 region of 26S rDNA, *C. flavescens* was approximately 0.015 and 0.031 substitution per base pair from the other two *Bulleromyces* species *C. aureus* and *C. laurentii*, respectively. Based on the same region, *C. flavescens* is much more distant (0.112 substitute per base pair) from the well studied opportunistic human pathogen *C. neoformans*, which is in the *Filobasidiella* clade (Fonseca et al., 2011).

*C. flavescens* OH182.9, the biocontrol strain against Fusarium head blight, was formerly reported as *C. nodaensis* (Khan et al., 2001). This was because OH182.9’s ITS1-5.8S-ITS2 sequence was identical to and its D1/D2 of 26S rDNA sequence was highly similar to *C. flavescens* G60, which was then reported as *C. nodaensis* (Khan et al., 2001; Sato et al., 1999; Sugita, Takashima, Ikeda, Nakase, & Shinoda, 2000). “C.
nodaensis” is a nomen invalidum since no Latin identification was reported (Sato et al., 1999; Takashima et al., 2003).

**Physiology**

The physiological characteristics of *C. flavescens* are similar to other *Cryptococcus* species in the *Bulleromyces* clade (Fonseca et al., 2011). *C. flavescens* was able to assimilate the following carbon sources: D-α-aminobutyric acid (DBB), L-arabinose, cellobiose, ethanol, galactitol, galactose, D-galacturonic acid, D-glucitol, D-gluconate, glucose, D-glucuronate, 2-keto-D-gluconate, lactose, maltose, D-mannitol, melibiose, melezitose, methyl-α-D-glucocide, *myo*-inositol, raffinose, L-rhamnose, D-ribose, salicin, succinate, sucrose, trehalose and D-xylose (Fonseca et al., 2011; Takashima et al., 2003). It was able to utilize erythritol and ribitol but assimilation was slow (Fonseca et al., 2011). It was not able to utilize *N*-acetyl-D-glucosamine, creatinine, hexadecane, inulin or methanol. It showed variable results in assimilating D-arabinose, citrate, D-glucosamine, glycerol, D,L-lactate, soluble starch and L-sorbose. It was unable to grow in 50% glucose solution and showed variable results for growing in 10% NaCl plus 5% glucose. It was able to grow at 25°C and 30°C but showed none to weak growth at 37°C. It was able to form starch and to produce urease (Fonseca et al., 2011; Kurtzman, 1973). It differs from *C. laurentii* in that it was not able to assimilate *N*-acetyl-D-glucosamine or creatinine, and that it was able to grow without amended vitamin, the latter of which is uncommon within the Tremellales order (Fonseca et al., 2011).

**Sexual reproduction**

Mating was observed among the *C. flavescens* strains isolated from the surfaces of dent corn kernels with various levels of moisture (Kurtzman, 1973). Mating test was performed by mixing the suspensions of each pair of strains and streaking the mixture onto 5% malt extract agar plates at 15°C. Formation of conjugation tube and occurrence of hypha outgrowth from mucoid portion of colony was considered as positive mating reaction. Greater mating responses occurred at 15°C than at 20 or 25°C and on malt extract agar than corn meal, yeast malt or V-8 agar. The two strains that were able to
mate were considered as being opposite mating types. Based on the mating test, strains NRRL Y-7139, Y-7372, Y-7374, Y-7375, Y-7376, Y-7379 were designated mating type α (alpha); Y-7371, Y-7373, Y-7377 mating type a (‘aye’). Y-7139 and Y-7371 were tested for mating with type strain CBS 942$^T$ = NRRL Y-1401$^T$, however, no evidence of mating was observed during prolonged incubation under the aforementioned test conditions.

Hyphae originated from either the conjunction at the tips of fused conjugation tubes or one of the two fused cells. Yeast-form cells were uninucleate; cells in hyphae near the fusion point of conjugation tubes were binucleate while those in further extended hyphae turned uninucleate. Clamp connections were not observed on either type of hyphae. Mycelium originating from mating was squashed and streaked on yeast malt agar to check whether segregating mating types can be obtained. Segregation of mating types was observed from the cross of Y-7139 (α) × Y-7371 (a) as mucoid colonies of each mating type, while mucoid colonies of mating type α only were obtained from the cross of Y-7376 (α) and Y-7371. However, there was no certainty that the segregating mating types resulted from meiosis (Kurtzman, 1973).

Thick-walled chlamydospore-like structures filled with lipid droplets, either intercalary or terminal, were observed on older portions of hyphae. Papilla was observed on the distal ends of many of such structures that are terminal. Most of such structures had one or two nuclei, three or four nuclei per spore were observed but rare (Kurtzman, 1973). Germination of these chlamydospore-like structures produced basidium-like structures (Fonseca et al., 2011; Kurtzman, 1973). If the latter were true basidia, the former would actually be teliospores, then *C. flavescens* would be a teliomorphic species (Fonseca et al., 2011; Kurtzman, 1973).

Need to understand more about the biology, ecology and diversity of *C. flavescens*
OH182.9_3C (3C) is a promising biocontrol agent against Fusarium head blight (FHB), and currently a biopesticide based on 3C is under development. To facilitate the development of this biopesticide, a better understanding of the biology, ecology and diversity of *C. flavescens* is necessary. Specifically, the following questions will be addressed in this thesis.

The biocontrol mechanism of 3C is unknown. By understanding the mechanism, better use of 3C as a biopesticide can be made through adjusting formulation and application practices. Identification of genes related to biocontrol of 3C will provide insight into its biocontrol mechanism.

To register 3C as a biopesticide, ecological risk assessment including environmental fate determination is required by US Environmental Protection Agency (http://www.epa.gov/pesticides/ecosystem/ecorisk.htm). Previous studies showed 3C was able to colonize and multiply on wheat (Schisler et al., 2009; Schisler et al., 2010). This means that after 3C is sprayed in the field, it may disperse from inoculated area to non-inoculated area and persist in the field and crop products even when no additional inoculation is performed. Therefore, the dispersal and persistence patterns of 3C throughout the crop production system need to be investigated to determine its environmental fate.

Multiple strains within the same species are known to have biocontrol activity (Pal & McSpadden Gardener, 2006). Other *C. flavescens* strains than 3C may also show biocontrol efficacy against FHB. Previous studies showed association of genetic markers with disease suppression and the success of using those markers to recover novel plant-beneficial bacteria (Benitez & McSpadden Gardener, 2009; Benítez et al., 2007). A better understanding of the diversity of and phylogenetic relationship among available *C. flavescens* strains will enable development of such markers to assist bio-prospecting for novel *C. flavescens* that are potential biocontrol agents.
<table>
<thead>
<tr>
<th>Wheat cultivar</th>
<th>Mixed&lt;sup&gt;b&lt;/sup&gt; variables</th>
<th>F&lt;sub&gt;g&lt;/sub&gt; applied&lt;sup&gt;c&lt;/sup&gt; vs. tested</th>
<th>Disease severity (%)</th>
<th>Disease incidence (%)</th>
<th>100 kernel weight (g)</th>
<th>Visually flawless kernel (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Norm</td>
<td>Z3639</td>
<td>Buffer 90 OH 182.9 39* ↓56.7</td>
<td>Buffer 95 OH 182.9 72* ↓24.2</td>
<td>Buffer 1.5 OH 182.9 3.0* ↑100.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>DOAM 180378</td>
<td>76 69 9.2</td>
<td>91 84 7.7</td>
<td>1.8 2.0* 11.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>Fg-9-96</td>
<td>54 51 5.6</td>
<td>66 65 1.5</td>
<td>3.2 3.5* 9.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Antagonist level</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>10% 4 h before</td>
<td>59 43* 27.1</td>
<td></td>
<td>1.8 1.9 5.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>10% Imme. before</td>
<td>86 64* 25.6</td>
<td></td>
<td>1.6 2.3* 43.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>10% Imme. after</td>
<td>81 60* 25.9</td>
<td></td>
<td>1.7 2.3* 35.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>10% 4 h after</td>
<td>85 45* 47.1</td>
<td></td>
<td>1.4 2.7* 92.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>50% 4 h before</td>
<td>59 43* 27.1</td>
<td></td>
<td>1.8 1.8 0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>50% Imme. before</td>
<td>86 49* 43.0</td>
<td></td>
<td>1.6 2.2* 37.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>50% Imme. after</td>
<td>81 60* 25.9</td>
<td></td>
<td>1.7 2.3* 35.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norm</td>
<td>50% 4 h after</td>
<td>85 58* 31.8</td>
<td></td>
<td>1.4 2.0* 42.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Media C:N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renville (durum)</td>
<td>6.5</td>
<td>80 50* 37.5</td>
<td>100 84 16.0</td>
<td>2.26 2.33 3.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renville (durum)</td>
<td>11</td>
<td>80 44* 45.0</td>
<td>100 84 16.0</td>
<td>2.26 2.66* 17.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben (durum)</td>
<td>6.5</td>
<td>82 71 13.4</td>
<td>100 97 3.0</td>
<td>3.28 2.61* -20.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben (durum)</td>
<td>11</td>
<td>82 76 7.3</td>
<td>100 94 6.0</td>
<td>3.28 2.70* -17.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renville (durum)</td>
<td>50</td>
<td>50 27* 46.0</td>
<td>96 79* 17.7</td>
<td>1.89 2.10* 11.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ben (durum)</td>
<td>46</td>
<td>46 43 6.5</td>
<td>97 87 10.3</td>
<td>3.19 2.74* -15.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Summary of greenhouse data on biocontrol efficacy of OH182.9 against Fusarium head blight<sup>a</sup>
<table>
<thead>
<tr>
<th>Location, year</th>
<th>Wheat cultivar</th>
<th>Mixed variables</th>
<th>Disease severity (%)</th>
<th>Disease incidence (%)</th>
<th>100 kernel weight (g)</th>
<th>DON (ppm)</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Buffer OH 182.9</td>
<td>% ↓</td>
<td>Buffer OH 182.9</td>
<td>% ↓</td>
<td>Buffer OH 182.9</td>
<td>% ↓</td>
</tr>
<tr>
<td><strong>Antagonist level</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wooster, OH, 1999</td>
<td>Pioneer 2545</td>
<td>10%</td>
<td>11 4.6* 58.2</td>
<td>34.4 23.1* 32.8</td>
<td>2.76 2.98 8.0</td>
<td>9.9 5.7 42.4</td>
<td>8</td>
</tr>
<tr>
<td>Wooster, OH, 1999</td>
<td>Pioneer 2545</td>
<td>50%</td>
<td>11 7.1* 35.5</td>
<td>34.4 27* 21.5</td>
<td>2.76 2.96 7.2</td>
<td>9.9 8.0 19.2</td>
<td>8</td>
</tr>
<tr>
<td>Wooster, OH, 1999</td>
<td>Freedom</td>
<td>10%</td>
<td>3.5 3.7 -5.7</td>
<td>17.8 16.1 9.6</td>
<td>2.89 2.84 -1.7</td>
<td>0.9 0.7 22.2</td>
<td>9</td>
</tr>
<tr>
<td>Wooster, OH, 1999</td>
<td>Freedom</td>
<td>50%</td>
<td>3.5 2.8 20.0</td>
<td>17.8 20.8 -16.9</td>
<td>2.89 2.94 1.7</td>
<td>0.9 0.8 11.1</td>
<td>9</td>
</tr>
<tr>
<td>Peoria, IL, 1999</td>
<td>Pioneer 2545</td>
<td>10%</td>
<td>2.0 0.9* 55.0</td>
<td>11.2 5.4* 51.8</td>
<td>3.3 3.4 3.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Peoria, IL, 1999</td>
<td>Pioneer 2545</td>
<td>50%</td>
<td>2.0 1.6 20.0</td>
<td>11.2 6.7* 40.2</td>
<td>3.3 3.4 3.0</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Peoria, IL, 1999</td>
<td>Freedom</td>
<td>10%</td>
<td>1.0 0.3* 70.0</td>
<td>8.3 3.3* 60.2</td>
<td>3.1 3.2 3.2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Peoria, IL, 1999</td>
<td>Freedom</td>
<td>50%</td>
<td>1.0 0.6 40.0</td>
<td>8.3 3.8* 54.2</td>
<td>3.1 3.0 -3.2</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td><strong>Media C:N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wooster, OH, 2000</td>
<td>Pioneer 2545</td>
<td>6.5</td>
<td>8.4 2.9* 65.5</td>
<td>40.0 21.4* 46.5</td>
<td>2.10 2.25 7.1</td>
<td>3.9 3.4 12.8</td>
<td>11</td>
</tr>
<tr>
<td>Wooster, OH, 2000</td>
<td>Pioneer 2545</td>
<td>11</td>
<td>8.4 3.4* 59.5</td>
<td>40.0 22.0* 45.0</td>
<td>2.10 2.25 7.1</td>
<td>3.9 2.5 35.9</td>
<td>11</td>
</tr>
<tr>
<td>Wooster, OH, 2000</td>
<td>Freedom</td>
<td>6.5</td>
<td>3.0 3.0 0.0</td>
<td>24.7 27.8 -12.6</td>
<td>2.26 2.17 -4.0</td>
<td>3.9 3.8 2.6</td>
<td>11</td>
</tr>
<tr>
<td>Wooster, OH, 2000</td>
<td>Freedom</td>
<td>11</td>
<td>3.0 2.7 10.0</td>
<td>24.7 25.3 -2.4</td>
<td>2.26 2.25 -0.4</td>
<td>3.9 3.4 12.8</td>
<td>11</td>
</tr>
<tr>
<td>Peoria, IL, 2000</td>
<td>Pioneer 2545</td>
<td>6.5</td>
<td>7.6 4.1* 46.1</td>
<td>32.1 27.1 15.6</td>
<td>2.57 2.53 -1.6</td>
<td>5.3 4.1 22.6</td>
<td>10</td>
</tr>
<tr>
<td>Peoria, IL, 2000</td>
<td>Pioneer 2545</td>
<td>11</td>
<td>7.6 3* 60.5</td>
<td>32.1 20.9* 34.9</td>
<td>2.57 2.56 -0.4</td>
<td>5.3 5.2 1.9</td>
<td>10</td>
</tr>
<tr>
<td>Peoria, IL, 2000</td>
<td>Freedom</td>
<td>6.5</td>
<td>0.6 0.9 -50.0</td>
<td>7.9 10.4 -31.6</td>
<td>2.71 2.81 3.7</td>
<td>2.2 1.7 22.7</td>
<td>10</td>
</tr>
<tr>
<td>Peoria, IL, 2000</td>
<td>Freedom</td>
<td>11</td>
<td>0.6 0.9 -50.0</td>
<td>7.9 11.7 -48.1</td>
<td>2.71 2.82 4.1</td>
<td>2.2 6.2 -181.8</td>
<td>10</td>
</tr>
<tr>
<td>Langdon, ND, 2000</td>
<td>Grandin</td>
<td>6.5</td>
<td>12.9 12.3 4.7</td>
<td>28.1 37.2* -32.4</td>
<td>3.57 3.42 -4.2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Langdon, ND, 2000</td>
<td>Russ</td>
<td>4.7</td>
<td>4.6 2.1</td>
<td>17 15.3 10.0</td>
<td>3.42 3.34 -2.3</td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Summary of field data on biocontrol efficacy of OH182.9 against Fusarium head blight

* Continued
Table 2. Continued

<table>
<thead>
<tr>
<th>Location, year</th>
<th>Wheat cultivar</th>
<th>Mixed b variables</th>
<th>Disease severity (%)</th>
<th>Disease incidence (%)</th>
<th>100 kernel weight (g)</th>
<th>DON (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Buffer OH 182.9</td>
<td>% ↓ d</td>
<td>Buffer OH 182.9</td>
<td>% ↓ d</td>
<td>Buffer OH 182.9 % ↓ d</td>
</tr>
<tr>
<td>Peoria, IL, 1999</td>
<td>Renville (durum)</td>
<td>1.9</td>
<td>1.3</td>
<td>31.6</td>
<td>8.8</td>
<td>11.8</td>
</tr>
<tr>
<td>Peoria, IL, 2000</td>
<td>Renville (durum)</td>
<td>14</td>
<td>6*</td>
<td>57.1</td>
<td>62</td>
<td>36*</td>
</tr>
<tr>
<td>Peoria, IL, 2000</td>
<td>Ben (durum)</td>
<td>10</td>
<td>9</td>
<td>10.0</td>
<td>55</td>
<td>43*</td>
</tr>
<tr>
<td>Langdon, ND, 2000</td>
<td>Renville (durum)</td>
<td>5.9</td>
<td>6.7</td>
<td>-13.6</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>Langdon, ND, 2000</td>
<td>Ben (durum)</td>
<td>4.7</td>
<td>3.7</td>
<td>21.3</td>
<td>14</td>
<td>11</td>
</tr>
</tbody>
</table>

Foot notes for both Table 1 and Table 2

a Values ending with an asterisk (*) were significantly different from buffer (negative control) in Fisher’s protected LSD test (α = 0.05); values without * are either not significant or statistics not reported. Empty cells indicate the predictor variable was not specified or the response variable was not measured.

b Multiple predictor variables are shown in this column, including pathogen isolates (*Fg* = *Fusarium graminearum*), OH182.9 inoculum (antagonist) concentration, and the carbon to nitrogen ratio of the media used to produce OH182.9 inoculum (media C:N).

c *Fusarium graminearum* inoculum applied 4 h before/after or immediately (imme.) before/after buffer (negative control) or OH182.9 inoculum

d Percentage reduction (↓) or increase (↑) by OH182.9 compared to buffer (negative control)

e Reference (Ref.):

1. Table 4 in (Khan et al., 1999a)
2. Table 2 in (Khan et al., 2001); Table 1 in (Khan et al., 1999b)
3. Table 4 & 5 in (Khan et al., 2001); Table 2 & 3 in (Khan et al., 1999b)
4. Table 4 & 5 in (Khan et al., 2001)
5. Figure 1 & 2 in (Khan et al., 2004); Table 2 in (Boehm et al., 1999)
6. Table 2 in (Khan et al., 2004); Table 3 in (Schisler et al., 2000)
7. Table 2 in (Khan et al., 2004)
8. Table 3 in (Khan et al., 2004); Table 3 in (Boehm et al., 1999)
9. Table 3 in (Khan et al., 2004)
10. Table 4 in (Khan et al., 2004); Table 2 in (Schisler et al., 2000)
11. Table 5 in (Khan et al., 2004); Table 2 in (Schisler et al., 2000)
12. Table 2 in (Schisler et al., 2002); Table 2 in (Khan et al., 1999a)
13. Table 2 in (Schisler et al., 2002); Table 3 in (Khan et al., 1999a)
14, 15, 16. Table 3, 4, 5, respectively, in (Schisler et al., 2002)
Chapter 2: Genomic Analysis of *C. flavescens* OH 182.9_3C

Abstract

The genome of *Cryptococcus flavescens* 3C was sequenced by high-throughput method and assembled into scaffolds. Draft annotations were obtained for the genome through Comparative Fungal Genome Platform, Augustus prediction, Integrative Microbial Genome at Joint Genome Institute and Blastx search against NCBI non-redundant protein sequences (nr). Nine conserved regions, including rDNA-ITS, seven protein-encoding genes and one mostly anonymous region, were identified from 3C genome. Primers were designed to amplify parts of these regions for a quantitative PCR assay to measure the abundance of 3C-like *C. flavescens* in the field and for multilocus sequence typing to characterize the genetic diversity among multiple *C. flavescens* strains. An attempt was also made to identify 3C genes potentially involved in biocontrol. Regions homologous to non-ribosomal peptide synthetase and glucanase, which were previously shown to be involved in biocontrol of other microorganisms, were identified in 3C genome. Whether they are actual biocontrol genes needs further investigation.

Introduction

Need for genetic markers to detect and characterize 3C and other *C. flavescens* strains

As will be detailed in later chapters, amplification targets for quantitative PCR to detect 3C in the field and for multilocus sequence typing are needed. The 3C genome will
provide resources for designing primers to amplify these targets. A draft genome
annotation is needed for narrowing down to conserved regions that are better candidates
for marker development than highly variable regions. Genome annotations can also be
used for deducing the biocontrol mechanisms of 3C.

Types of biocontrol mechanism in general and yeast biocontrol agents for plant
diseases

Biological control agents (BCAs) can reduce disease damages in plants by suppressing pathogens or inducing host resistance. BCAs can suppress plant pathogens through production of deleterious compounds, hyperparasitism/predation, competition for resources and/or interference without clear purposes (Pal & McSpadden Gardener, 2006). Antibiotics were known to be produced by various bacteria, such as *Pseudomonas* spp. (2,4-diacetylphloroglucinol, phenazine, pyoluteorin, pyrrolnitrin), *Bacillus subtilis* (cyclic lipopeptide), *Agrobacterium radiobacter* (Agrocin 84) and *Pantoea agglomerans* (herbicolin). Such antibiotics can suppress plant pathogens such as *Pythium* spp. (damping off of sugar beet, cotton and tomato), *Botrytis cinerea* (gray mold of tomato), *Rhizoctonia solani* (damping off of eggplant), *Gaeumannomyces graminis var. tritici* (take-all of wheat) and *Erwinia amylovora* (fire blight of apple and pear). Unregulated wastes of BCAs, such as hydrogen cyanide, carbon dioxide and ammonia, may also suppress plant pathogens. BCAs can compete with pathogens for resources through exudates/leachates consumption, siderophore scavenging and physical niche occupation. Pathogens may also be interfered by BCAs that block soil pores, consume germination signals and/or confuse pathogen signaling pathways by molecular cross-talk (Pal & McSpadden Gardener, 2006).

Induction of host resistance is believed to be based on the basal defense mechanism of host against general microorganisms. Detection of microbial presence will trigger host metabolic pathways that prepare host to fight against potentially harmful microbial invasion. However, BCAs can trigger those pathways without doing actual harm to the host, instead, they will enable the host to defend itself against future pathogen attack more effectively. The compounds produced by BCAs that can induce host resistance
include enzymes (peroxidase, chitinase, β-1,3-glucanase), lipopolysaccharide, siderophore, 2,4-diacetylphloroglucinol, 2,3-butanediol, Z,3-hexenal and iron-regulated factors (Pal & McSpadden Gardener, 2006). Most of these compounds trigger the induced systemic resistance (ISR) pathway mediated by jasmonic acid and/or ethylene. But some of them, such as siderophore, can also trigger the systemic acquired resistance (SAR) pathway mediated by salicylic acid, which is more often triggered by pathogen and insect attack. The following bacterial BCAs have been shown to induce host resistance: *Bacillus mycoides*, *B. pumilus*, *Pseudomonas fluorescens*, *Pseudomonas putida* and *Serratia marcescens* (Pal & McSpadden Gardener, 2006).

*Trichoderma* spp. are by far the most extensively studied fungal BCAs (Howell, 2003). Mycoparasitism is well known for *Trichoderma*, which can penetrate the mycelium of *R. solani* and form haustoria. *T. virens* can produce gliotoxin, an antibiotic suppressing *R. solani* and *Pythium ultimum*. *Trichoderma* spp. showed rapid colonization of roots and thus may compete with plant pathogens for rhizosphere space and nutrients. They also produce lytic enzymes such as chitinases and glucanases, which may degrade cell wall components of plant pathogens and form elicitors to trigger plant defense mechanisms. In addition, *T. virens* was able to metabolize the compounds secreted by cotton seed which stimulate germination of damping-off pathogens *P. ultimum* and *Rhizopus oryzae*.

Most of the yeast BCAs studied so far are against postharvest diseases. A strain of *Cryptococcus laurentii*, a species closely related to *C. flavescens*, showed significant control of pear fruit decay caused by gray mold (*Botrytis cinerea*) and blue mold (*Penicillium expansum*) (Benbow & Sugar, 1999). *Rhodosporidium paludigenum*, a Basidiomycete yeast in a different order than *Cryptococcus*, reduced incidence of postharvest gray mold on cherry tomato possibly through competition for nutrients (Wang et al., 2010). Aspire, a commercial biocontrol product based on the Ascomycete yeast *Candida oleophila*, has been registered in United States and Israel for controlling postharvest decay of citrus caused by *Penicillium digitatum*, *P. italicum* and *Geotrichum candidum* (Droby et al., 1998). A combination of ultraviolet light C and *Debaryomyces hansenii*, an Ascomycete yeast closely related to the model yeast *Saccharomyces*, showed
comparable efficacy to commercial fungicide against brown rot of peach caused by *Monilinia fructicola* (Stevens et al., 1997).

The molecular mechanisms of how yeast BCAs control plant diseases have not been extensively studied. Transcriptomic change in grape fruit peel in response to an Ascomycete yeast BCA *Metschnikowia fructicola* was reported (Hershkovitz et al., 2012). Genes upregulated by the application of *M. fructicola* included Respiratory burst oxidase (*Rbo*), chitinase (*CHI*), phenylalanine ammonia-lyase (*PAL*), chalcone synthase (*CHS*) and 4-coumarate-CoA ligase (*4CL*) and signaling related genes such as mitogen-activated protein kinase (*MAPK*), mitogen-activated protein kinase kinase (*MAPKK*) and G-proteins. The genes in the scavenging/antioxidant systems such as peroxidase (*POD*), superoxide dismutase (*SOD*) and catalase (*CAT*) were downregulated. Moreover, the downregulation of those genes was correlated with a higher level of hydrogen peroxide, superoxide anion and hydroxyl radical production in surface wounds of grape fruit treated with *M. fructicola*. This indicates that the yeast BCA triggered an oxidative burst in wound tissue of host and thus led to activation of host defense responses against postharvest pathogens (Hershkovitz et al., 2012). However, the gene composition and expression of yeasts related to biocontrol have not been well characterized.

**Genomic analysis on C. flavescens OH182.9_3C (3C) can help to decipher its biocontrol mechanisms**

Unlike most of the yeast biocontrol agents (BCAs) that have been studied, 3C colonizes and controls disease on wheat heads preharvest (Khan et al., 2001; Schisler et al., 2009; Schisler et al., 2010) instead of on juicy fruits postharvest. This suggests that high level of similarity between the biocontrol mechanisms of 3C and those postharvest yeast BCAs cannot be assumed. Another well studied biocontrol fungus is *Trichoderma* spp., which is a filamentous ascomycete that covers a wide range of biocontrol mechanisms at rhizosphere. *Trichoderma* spp. have been reported as being capable of mycoparasitism, antibiotic production, lytic enzyme production, competition for rhizosphere resources, metabolism of germination stimulants and host resistance induction (Howell, 2003). First of all, 3C is a yeast with very limited filamentous stage in
its life cycle and such stage has not been observed on wheat (Kurtzman, 1973). Second, again, the biocontrol-relevant niche of 3C is wheat head instead of rhizosphere. This suggested that the *Trichoderma* model of biocontrol mechanism is not very applicable to 3C either. Therefore, alternative approaches to using similar BCAs as references are needed to investigating the biocontrol mechanism of 3C.

Genomic approach is promising for understanding the mechanisms of various biological activities since it provides a global view of the genetic features that encode and direct those activities (Rong, Baysal Gurel, Meulia, & McSpadden Gardener, 2012a; Rong, Baysal Gurel, Meulia, & McSpadden Gardener, 2012b). A draft genome of 3C is available under accession numbers CAUG01000001 to CAUG01000712 (X. Rong and B. B. McSpadden Gardener, submitted for publication). Annotating the genome will enable identification of 3C genes homologous to the known genes from other BCAs that are related to biocontrol, and in turn, deduction of possible biocontrol mechanisms of 3C based on the identified genes. The genome annotation can also serve as resources for developing markers to detect and characterize 3C and other *C. flavescens* strains. In reality, marker development was conducted prior to biocontrol-related gene identification, due to the prioritizing to achieve grant goals and that the former requires less curation in genome-wise annotation.

**Identification of genetic markers to detect and characterize 3C and other *C. flavescens* strains**

**Genome sequencing and assembly of 3C**

The genomic DNA of *Cryptococcus flavescens* OH182.9_3C was isolated using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA). The library for Illumina sequencing was prepared from the DNA fraction of ~ 300 bp using Illumina Paired-End Sample Preparation Kits, following the protocols provided by the manufacturer. This library was sequenced in one lane of a standard flow cell on Illumina
Genome Analyzer II (Illumina, San Diego, CA, USA) for 76 cycles, generating 46,300,240 of paired-end reads most of which were 75 nt long, which were equal to ~3.50 billion nucleotides. The reads that are shorter than 20 nt and/or have nucleotides with lower Sanger quality scores than 20 were removed, leaving 44,605,496 reads (95.72% of total).

The high quality reads were assembled in Velvet v1.0 (Zerbino & Birney, 2008; Zerbino, McEwen, Margulies, & Birney, 2009) using hash length (k-mer) = 31 nt and coverage cutoff = 5, resulting in an assembly with 22,830,355 nt in 3,580 scaffolds, N50 = 71,386 nt, max scaffold = 374,074 nt, using 41,362,769 reads. Short scaffolds were removed from the assembly using a threshold of minimum length 300 nt, leaving 712 scaffolds. This assembly had a total of 22.79 million nt and a median coverage of 45 ×. The sequences were deposited in GenBank/EMBL/DDBJ under accession numbers CAUG01000001 to CAUG01000712 (see Appendix A for the computer scripts used to process the sequences for submission).

**Identification of conserved regions that can be used as genetic markers**

Blastn search was performed on the Comparative Fungal Genome Platform (http://cfgp.riceblast.snu.ac.kr) using the 712 scaffolds of 3C as batch query against the NCBI Nucleotide collection (nr/nt) updated in mid-2008. Conserved regions that potentially can be used as markers were identified by searching the annotations of the hits of E-values ≥ 1e-10 for the following keywords: internal transcribe spacer or ITS, beta tubulin or btub, chitin synthase and heat shock protein 70 kDa or hsp70. For each keyword hit, blastx search querying the respective 3C sequence against the non-redundant protein sequences (nr) was conducted through NCBI web interface (http://blast.ncbi.nlm.nih.gov) in December, 2010. This was to confirm the 3C sequence was truly homologous to the known genes of respective keyword functions. Nine regions of the following putative functions were identified from 3C genome: internal transcribed spacer 1 and 2 plus 5.8S ribosomal RNA gene (ITS1-5.8S-ITS2), one β-tubulin gene (btub), one elongation factor 1 gene (EF1), three chitin synthase genes and three Heat shock protein 70 kDa genes (Hsp70) (Table 3).
The contig containing the ITS hit from batch blastn (CAUG01000137, NODE_197_length_2341_cov_1107.168335) was used as query for megablast search against the NCBI Nucleotide collection (nr/nt) as of May, 2013. The majority of the hits were the ITS1-5.8S-ITS2 region of various *C. flavescens* strains. The annotations of the hits were used to define the boundaries of ITS1, 5.8S rDNA and ITS2 in 3C. Augustus gene prediction was performed on scaffolds containing the aforementioned protein encoding gene hits from batch blastn (Table 3) on the web interface as of December, 2010 ([http://bioinf.uni-greifswald.de/augustus/](http://bioinf.uni-greifswald.de/augustus/)), using *C. neoformans* as model (Stanke, Steinkamp, Waack, & Morgenstern, 2004). The prediction output was used to define the boundaries of exons and introns in those genes. The score of Augustus prediction, which indicates the confidence for each feature, ranged from 0.22 to 1.00, where the majority was above 0.80 (Table 3). Discontiguous megablast search was conducted querying the 8 protein-encoding genes of 3C against the genome of *C. neoformans* JEC21 (AE017341 to AE017356) to confirm the Augustus prediction of genes and exon/intron boundaries. Hits in JEC21 were identified showing sequence homology of 50% to 72% at DNA level and approximate exon/intron match to each of the eight 3C genes.

**Primer design for quantitative PCR to detect 3C in the field**

A quantitative PCR assay was needed for quantification of 3C population in the field. The amplification target needs to be conserved enough so that polymorphism will not develop soon among the individuals of the 3C strain. Seven out of the nine conserved regions identified as above were subject to primer design by the Primer3 plug-in in Geneious version 5.3 ([Biomatters, http://www.geneious.com](http://www.geneious.com)). Primers were not designed for the Chitin synthase 7 region from CAUG01000482 and the Chitin synthase 1 region from CAUG01000683 (Table 3). The following restrictions were applied in Primer 3: product size 100 to 150 bp; primer size min 18, optimal 22, max 25; Tm min 57, optimal 60, max 72; %GC min 40, optimal 50, max 60; max Tm difference 100, max hairpin score 5, max poly-X 3, GC clamp 0, max primer dimer score 2, max 3’ stability 9. For each of the 7 regions, 10 pairs of primers were generated by
Primer3 and two pairs with the highest scores given by Primer3 and/or representing non-overlapping amplicons were chosen for wet-lab test (Table 4).

Genomic DNA of 3C and 6 other *C. flavescens* strains (NRRL Y-7372, Y-7373, YB-328, YB-601, YB-602 and YB-744) was used as templates to test the specificity of the 14 primer pairs in quantitative PCR. Y-7373, YB-601 and YB-602 were of the same genotype as 3C (3C genotype) while Y-7372, YB-328 and YB-744 were another genotype (non-3C genotype) (detailed in Chapter 4). One primer pair h31.2 amplifying part of the heat shock protein 70 kDa gene on CAUG01000237 was able to distinguish the two genotypes by a difference of approximately 14 in Threshold cycle (Ct) value. The two primer pairs amplifying parts of the β-tubulin (btub.1 and btub.2) distinguished the two genotypes by 11 to 12 in Ct value. Four other primer pairs targeting elongation factor 1 and other heat shock protein 70 genes (EF1.2, h22.1, h30.2, h31.1) showed slight differences in Ct (3, 3, 1.5, 2.5). The rest of the primers were not able to differentiate the two genotypes. The primer pair h31.2 was chosen for use in qPCR for field quantification of 3C since it showed highest differentiating power between the two genotypes among all the primers. To achieve the same level of qPCR detection by h31.2, the template concentration for non-3C genotype needs to be more than $10^4$ times of that for 3C genotype.

**Primer design for multilocus sequence typing to characterize *C. flavescens* strains**

Multilocus sequence typing was needed to understand the phylogenetic relationship among different *C. flavescens* strains. The sequence typing loci need to be conserved in the primer binding sites but polymorphic in the rest of the amplified regions so that the targets can be amplified from all the strains to be characterized while still be phylogenetically informative. The classic ITS5 and ITS4 primers widely used in fungal identification (White et al. 1990) were used to amplify the ITS1-5.8S-ITS2 region. The 8 protein encoding genes identified from 3C genome were subject to primer design by the Primer3 (Rozen & Skaletsky, 2000) plug-in in Geneious version 5.4 (Biomatters, [http://www.geneious.com](http://www.geneious.com)). The following restrictions were applied in Primer 3: product size 300 to 900 bp; primer size min 23, optimal 25, max 27; the other parameters were
the same as designing qPCR primers. For each of the 8 genes, 5 pairs of primers were generated by Primer3 and two pairs with the highest scores given by Primer3, including maximum amount of intron sequences and/or representing non-overlapping amplicons were chosen for wet-lab test (Table 5). For some genes, one forward/reverse primer and two reverse/forward primers were chosen, but this still gave rise to two different amplicons (Table 5).

Genomic DNA of 3C and 3 other *C. flavescens* strains (NRRL Y-7372, Y-7373 and YB-328) was used as templates to test whether the 17 pairs of primers can amplify from both genotypes. Y-7373 was of the same genotype as 3C (3C genotype) while Y-7372 and YB-328 were another genotype (non-3C genotype) (detailed in Chapter 4). Two primer pairs CS24.m1 (chitin synthase 1) and EF1.m1 (elongation factor 1) were able to amplify from 3C genotype but not from non-3C genotype. Primer pair CS10.m1 was able to amplify from 3C genotype and YB-328 but not from Y-7372. Primer pair CS22.m2 were able to amplify both genotypes but the amplicon length differed by genotype and the length difference was visible on agarose gel. The rest of the primer pairs were all able to amplify from both genotypes and no length difference was visible on gel. Nine pairs of primers were chosen for multilocus sequence typing based on ability to amplify both genotypes, length of intron in amplicon (longer preferred), band darkness (darker preferred for similar amplicon length) and level of non-target amplification (lower preferred) (Table 5).

**Genome-wide annotation for 3C through JGI, Augustus and blastx**

The genome annotation of 3C obtained through batch blastn search on CFGP was rough and not up to date. It was sufficient for marker development but not for identification of 3C genes related to biocontrol. Therefore, an improved annotation was obtained through a combination of JGI automatic annotation, Augustus and blastx.

The assembly of 3C was uploaded to the automated annotation platform Integrated Microbial Genomes at DOE Joint Genome Institute (JGI). Even though the taxonomic
placement of 3C was specified as eukaryote *Cryptococcus*, it seemed that the platform treated the assembly as a prokaryotic genome since the apparent exons of the same genes were predicted as separate genes. Therefore the annotations from JGI were far away from conclusive. JGI predicted 7,226 coding sequence (CDS), 71 tRNA, 7 rRNA and 6 miscellaneous RNA features in 3C genome, where 482 scaffolds contained features while the other 230 scaffolds did not. The 7 rRNA features were termed as 16S or 23S, thus seemed to be considered as bacterial rRNA by JGI.

The stand-alone command-line version of Augustus 2.6.1 was used to predict the boundaries of gene, exons and introns in the assembly of 3C (Stanke et al., 2004). *Cryptococcus neoformans* var. *neoformans* JEC21 was used as model. Here are the key prediction parameters shown as part of the command used: “augustus --species=cryptococcus_neoformans_neoformans_JEC21 --strand=both --singlestrand=false --genemodel=partial --codingseq=on --sample=100 --keep_viterbi=true --alternatives-from-sampling=true --minexonintronprob=0.08 --minmeanexonintronprob=0.3 --maxtracks=20 --exonnames=on”. Augustus predicted 9,063 gene, 13,454 transcript and 72,066 CDS features in 3C genome, where 589 scaffolds contained features while the other 123 scaffolds did not. The numbers of transcript and CDS features are much greater than the number of genes since alternative transcripts and CDSs were predicted and that the CDSs of each gene were predicted as separate features instead of a joint feature. The finished genome annotation of *Cryptococcus neoformans* JEC21 (accession AE017341 to AE017353 and AE017356), included 6,322 gene, 6,301 mRNA and 6,301 CDS features. The finished genome annotation of *Cryptococcus gattii* WM276 (accession CP000286 to CP000299), included 6,580 gene, 6,565 mRNA and 6,580 CDS features. Many more genes were predicted in 3C than in *C. neoformans* and *C. gattii* possibly because many gene fragments on short scaffolds were considered as individual genes by Augustus.

Batch blastx search querying the whole assembly of 3C against the NCBI non-redundant protein sequences (nr) was conducted through the bioinformatics platform Galaxy customized by the Molecular and Cellular Imaging Center at Ohio Agricultural Research and Development Center at Wooster, Ohio (galaxy.oardc.ohio-state.edu).
BLAST search results were retrieved as .xml and converted into tab-separated values using a customized version of Biopython script originally copyrighted to Wibowo Arindrarto (Appendix A). A total of 178,518 Blastx hits with E-value < 1e-100 were found on 404 scaffolds while no hit on the other 308 scaffolds.

Among the three annotation methods, Augustus predicted features on the largest number of scaffolds where no feature or hit was identified by the other two methods (Figure 1). In addition, no scaffold where both JGI and Blastx found feature/hit was predicted as no feature by Augustus. This indicates that Augustus is the most sensitive method in finding genes among the three. This is probably because Augustus is an *ab initio* gene caller, while the other two methods are partially or completely based on sequence homology. Since *C. flavescens* is highly distinct from the *Cryptococcus* species of which genome annotation is available, some genes predicted by Augustus may not be recognized by the other two methods due to low sequence similarity to database sequences. This also explains why no scaffold contained Blastx hits but not Augustus- or JGI-predicted features. Features/hits were found by all the three methods on 400 scaffolds, most of which were longer than 10,000 bp. No feature/hit was identified by any of the three methods for 103 scaffolds, most of which were shorter than 1500 bp. This is probably because such short scaffolds usually contain truncated genes. These gene fragments may be missed by *ab initio* method due to absence of mandatory gene components such as start/stop codon, and missed by homology-based method due to shortening of query sequence homologous to database sequences.

Non-ribosomal peptide synthetase (NRPS) was previously shown to participate in biosynthesis of antimicrobial peptibols in *Trichoderma* spp. (Wei, Yang, & Straney, 2005) and in siderophore biosynthesis of Ascomycetes (Oide et al., 2006). Blastx hits of apparent NRPS were found in four scaffolds of 3C genome (Table 6). One Augustus-predicted gene was found on each scaffold covering the query regions of those Blastx hits. JGI also predicted NRPS-related features within those regions. For the two Augustus-predicted genes on scaffolds CAUG01000141 and CAUG01000321, respectively, most of the Blastx hits were NRPS. The query-hit sequence identity was lower than 50% for the CAUG01000141 gene while higher than 70% for the
CAUG01000321 gene, indicating the latter was more likely to be true NRPS than the former was. Interestingly, no hits from Cryptococcus were identified for these two genes. The possible explanations of this include: (i) 3C acquired these genes from other lineages than those of the Cryptococcus species whose genomic sequences are publicly available (mainly human-pathogenic C. neoformans and C. gattii); (ii) those other Cryptococcus spp. had these genes in the past but then lost them; (iii) these genes arose from gene duplication events, the two copies underwent substantial divergence, then the copy more homologous to other Cryptococcus spp. was lost. Which hypothesis is really the case can be investigated by examining codon and nucleotide composition bias and by reconstructing the genealogy of these loci that involves enough taxa where the Blastx hits belong and more Cryptococcus species. For the other two Augustus-predicted genes, respectively from scaffolds CAUG01000126 and CAUG01000531, the best Blastx hits were from Cryptococcus but most hits were (L-)aminoadipate-semialdehyde dehydrogenase instead of NRPS. Since (L-)aminoadipate-semialdehyde dehydrogenase is known to be involved in lysine biosynthesis and does not seem connected to non-ribosomal peptide synthesis (http://www.uniprot.org/uniprot/P07702), these two 3C genes may not really be NRPS genes.

β-1,3-glucan is a major component of fungal cell wall (Bartnicki-Garcia, 1968), therefore, glucanases produced by biocontrol agents may contribute to suppression of fungal plant pathogens. β-1,3-glucanase is also known for induction of host resistance (Pal & McSpadden Gardener, 2006). Blastx hits of apparent glucanases were found in four scaffolds of 3C genome (Table 7). One Augustus-predicted gene was found on each scaffold covering the query regions of those Blastx hits. For the gene from scaffold CAUG01000468, many Blastx hits were exo-beta-1,3-glucanase, suggesting that this gene can potentially compromise the integrity of fungal cell wall. For the other three Augustus-predicted genes, most of the top Blastx hits were glycosyl hydrolases, which are not necessarily glucanases. The JGI-predicted features in these regions were no closer to glucanases than the Blastx hits.

Whether these apparent NRPS and glucanase genes contribute to biocontrol of 3C through production of antibiotic/siderophore, degrading cell wall of fungal pathogens or
inducing host resistance cannot be determined solely based on \textit{in silico} gene prediction. Whether those regions actually encode functional genes needs to be verified through transcriptional analysis. Gene knockout can be performed to find out whether the loss of these genes affect biocontrol efficacy of 3C. Details of how such work would be performed are described in Chapter 5.
<table>
<thead>
<tr>
<th>Putative function</th>
<th>Accession No.</th>
<th>Scaffold</th>
<th>Coordinate</th>
<th>Exon, Intron&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Number</td>
</tr>
<tr>
<td>ITS1-5.8S-ITS2</td>
<td>CAUG01000137</td>
<td>NODE_197_length_2341_cov_1107.168335</td>
<td>1,314 to 1,866</td>
<td>3, 2</td>
</tr>
<tr>
<td>β-tubulin</td>
<td>CAUG01000343</td>
<td>NODE_610_length_35550_cov_43.064079</td>
<td>19,021 to 20,661</td>
<td>6, 5</td>
</tr>
<tr>
<td>Chitin synthase 7</td>
<td>CAUG01000482</td>
<td>NODE_1075_length_225642_cov_46.200565</td>
<td>10,057 to 6,878</td>
<td>6, 5</td>
</tr>
<tr>
<td>Chitin synthase 5</td>
<td>CAUG01000018</td>
<td>NODE_22_length_83737_cov_46.708168</td>
<td>2,328 to 9,916</td>
<td>11, 10</td>
</tr>
<tr>
<td>Chitin synthase 1</td>
<td>CAUG01000683</td>
<td>NODE_2442_length_249147_cov_46.019089</td>
<td>38,212 to 42,048</td>
<td>4, 3</td>
</tr>
<tr>
<td>Elongation factor 1</td>
<td>CAUG01000040</td>
<td>NODE_50_length_28733_cov_43.824348</td>
<td>1,706 to 3,714</td>
<td>9, 8</td>
</tr>
<tr>
<td>Heat shock protein 70 kDa</td>
<td>CAUG01000166</td>
<td>NODE_227_length_115551_cov_47.223415</td>
<td>33,351 to 35,611</td>
<td>6, 5</td>
</tr>
<tr>
<td></td>
<td>CAUG01000224</td>
<td>NODE_301_length_52775_cov_47.424294</td>
<td>42,101 to 44,177</td>
<td>5, 4</td>
</tr>
<tr>
<td></td>
<td>CAUG01000237</td>
<td>NODE_318_length_195246_cov_45.768410</td>
<td>173,328 to 170,981</td>
<td>7, 6</td>
</tr>
</tbody>
</table>

<sup>a</sup>The rDNA (18S, 5.8S, 25S) and ITS regions were counted here as exon and intron, respectively.

<sup>b</sup>The lengths of exons were underlined while those of introns were not. All lengths were listed in order from the start to the end of each gene.

<sup>c</sup>The exons predicted with confidence score lower than 0.80 by Augustus. This is not applicable to the ITS1-5.8S-ITS2 region since prediction was done by megablast search against public database instead of Augustus.

Table 3. The conserved regions of 3C genome used for marker development
<table>
<thead>
<tr>
<th>Target gene</th>
<th>Scaffold accession #, in original assembly</th>
<th>Relative to Exon (E) and/or Intron (I)</th>
<th>Amplicon span</th>
<th>Primer location</th>
<th>Primer name</th>
<th>Primer sequence (5’ -&gt; 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>β-tubulin</td>
<td>CAUG010000343, NODE_610</td>
<td>I -&gt; E, 102 bp</td>
<td>I</td>
<td>E</td>
<td>btub.1_F</td>
<td>TGCTTCCACGCCTGATGATGCCAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I -&gt; E, 134 bp</td>
<td>E</td>
<td>I</td>
<td>btub.1_R</td>
<td>TGTACGAGCCCTGGCCTGGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>btub.2_F</td>
<td>TGGTGCACCGGGTTCATCAATAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>btub.2_R</td>
<td>AACCCCTGGAGGCAGTCCGCTT</td>
</tr>
<tr>
<td>Chitin synthase 5</td>
<td>CAUG01000018, NODE_22</td>
<td>E, 145 bp</td>
<td>E</td>
<td>E</td>
<td>CS.1_F</td>
<td>TCAAACGCCTCCGGCCCTCAAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>E, 138 bp</td>
<td>E</td>
<td>E</td>
<td>CS.1_R</td>
<td>GCCTGAAGCCGGATTTGCGAT</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>E</td>
<td>CS.2_F</td>
<td>GCGTGGCGGGAGAAATTCGAC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>E</td>
<td>CS.2_R</td>
<td>GTGTGACTGCAAGTCCGCA</td>
</tr>
<tr>
<td>Elongation factor 1</td>
<td>CAUG01000040, NODE_50</td>
<td>E, 112 bp</td>
<td>E</td>
<td>E</td>
<td>EF1.1_F</td>
<td>TGAGGCGCCCATTTGTAAGCTC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>E</td>
<td>EF1.1_R</td>
<td>GACAAGGTTCTGTGCAATGCTT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>I -&gt; E -&gt; I, 115 bp</td>
<td>I</td>
<td>E -&gt; I</td>
<td>EF1.2_F</td>
<td>TCGTTGACTCTTTGAGACG</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>E -&gt; I</td>
<td>EF1.2_R</td>
<td>AAGGCGGCACTAAACTGCTGT</td>
</tr>
<tr>
<td>Heat shock protein 70 kDa</td>
<td>CAUG01000166, NODE_227</td>
<td>I (1b) -&gt; E, 146 bp</td>
<td>I (1b) -&gt; E</td>
<td>h22.1_F</td>
<td>GACTTCAACGACGGCAGCGAT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I -&gt; E -&gt; I, 146 bp</td>
<td>E</td>
<td>h22.1_R</td>
<td>ACGAGTACTGCGCTGGCCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>h22.2_F</td>
<td>AGAGTCACTCCGGCAGTGCTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>h22.2_R</td>
<td>TCTTTGACTCTCAAGGCGGAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAUG01000224, NODE_301</td>
<td>E -&gt; I -&gt; E, 120 bp</td>
<td>E</td>
<td>h30.1_F</td>
<td>AGGGCGAGCCAGTGTGCTGTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>h30.1_R</td>
<td>TCGGAGACCGTGCTGCTGCG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>h30.2_F</td>
<td>ACCAAAGCAGCTGTGCTGAG</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>h30.2_R</td>
<td>CGTTGACTCTGCTCGCTCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CAUG01000237, NODE_318</td>
<td>I -&gt; E -&gt; I, 114 bp</td>
<td>E</td>
<td>h31.1_F</td>
<td>CGGTCTTACCCCTTGGAGGCAA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>E</td>
<td>h31.1_R</td>
<td>GCAAGGTGCAACAGCTGAGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I -&gt; E</td>
<td>I</td>
<td>h31.2_F</td>
<td>CGTCGACGTGTGGCAGTGCTA</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>I -&gt; E</td>
<td>I</td>
<td>h31.2_R</td>
<td>GCTGCTTCTTTGCGGCTGTA</td>
<td></td>
</tr>
<tr>
<td>5.8S-ITS2</td>
<td>CAUG01000137, NODE_197</td>
<td>5.8S -&gt; ITS2, 120 bp</td>
<td>5.8S</td>
<td>5I.1_F</td>
<td>TCTTTGACGACCTTGGCGCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5.8S</td>
<td>5I.1_R</td>
<td>TGGTGGACAGACCCAACTGCA</td>
<td></td>
</tr>
<tr>
<td>ITS2</td>
<td>CAUG01000137, NODE_197</td>
<td>ITS2, 136 bp</td>
<td>ITS2</td>
<td>I2.1_F</td>
<td>AGGGCGTGGACTTTGCGTGCT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ITS2</td>
<td>I2.1_R</td>
<td>AAGGTCGCGAGTGGCAGTT</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The primers designed for quantitative PCR to detect 3C in field
<table>
<thead>
<tr>
<th>Scaffold accession #, in original assembly; target gene</th>
<th>Primer names</th>
<th>Sequence 5' -&gt; 3'</th>
<th>Amplicon size (bp)</th>
<th>Chosen for MLST</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAUG01000137, NODE_197; ITS1-5.8S-ITS2</td>
<td>ITS5</td>
<td>GGAAGTAAAGTCGTAACAAGG</td>
<td>553</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>ITS4</td>
<td>TCCTCGCTTTATTGATATGC</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>CAUG01000343, NODE_601; β-tubulin</td>
<td>btb.m1-2F</td>
<td>GGGCGGTCTTAGATTGACTTGGAAACC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>btb.m1R</td>
<td>GCTTGAGGTTTCGGAAGCAATGTC</td>
<td>620</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>btb.m2R</td>
<td>TCCTTCGATACAGAAGTGTACCGT</td>
<td>585</td>
<td>Yes</td>
</tr>
<tr>
<td>CAUG01000482, NODE_1075; chitin synthase 7</td>
<td>CS10.m1F</td>
<td>TATCCGCTCAACCATACCCGACAG</td>
<td>666</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>CS10.m1R</td>
<td>GTTCACCCCTAGGTCCAGGCGTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CS10.m2F</td>
<td>GATCCACTTCTCCGCTGTACCAGA</td>
<td>420</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CS10.m2R</td>
<td>CTATGCGGGCGTAGACCAGGAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAUG01000018, NODE_22; chitin synthase 5</td>
<td>CS22.m1F</td>
<td>CCAGCTCTGTGGTTTCTGCTGCGA</td>
<td>677</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>CS22.m1R</td>
<td>CGCAGGGCCGTACTCGTTCTTCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CS22.m2F</td>
<td>AAGAAGGGAGTGAGGAAGCAGT</td>
<td>868</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CS22.m2R</td>
<td>ACTTTGCTGTTCTGTTGAGACG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAUG01000683, NODE_2442; chitin synthase 1</td>
<td>CS24.m1F</td>
<td>TTGTGCTGGATTGCGTGTCTGATG</td>
<td>423</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>CS24.m1R</td>
<td>GACTCTTGTGTGAGGGACTGCGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>CS24.m2F</td>
<td>ACAGGCTTGGAGGGAGAAGTACAG</td>
<td>356</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>CS24.m2R</td>
<td>CAAAGCGAGGTACGTCGATG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAUG01000040, NODE_50; elongation factor 1</td>
<td>EF1.m1F</td>
<td>AGGAGACCTCAAACATGCCGTGTA</td>
<td>545</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>EF1.m1R</td>
<td>ACCTTTTCCTGTGCAGATCGCTTGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>EF1.m2F</td>
<td>CGTGGTACAAGGGAGTGAGCAAGG</td>
<td>658</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>EF1.m2R</td>
<td>CAATGTGGCGAGCTACGTCGAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAUG01000166, NODE_227; heat shock protein 70 kDa</td>
<td>h22.m1F</td>
<td>TCGATCCAAAGCTCACCCTCTCTTGC</td>
<td>862</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>h22.m1-2R</td>
<td>ATCGACAGAGCGAAACATCGACG</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>h22.m2F</td>
<td>CTCGTCGCCATATCGTTTACCTC</td>
<td>831</td>
<td>Yes</td>
</tr>
<tr>
<td>CAUG01000224, NODE_301; heat shock protein 70 kDa</td>
<td>h30.m1-2F</td>
<td>TCCGCTAGATGCTCATCGAGGAGG</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>h30.m1R</td>
<td>AGTTCGAGATGAGGACATCGTTTCG</td>
<td>834</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>h30.m2R</td>
<td>CGGAGTTCATGCGAAGTGTTCCTTG</td>
<td>375</td>
<td>No</td>
</tr>
<tr>
<td>CAUG01000237, NODE_318; heat shock protein 70 kDa</td>
<td>h31.m1F</td>
<td>ATGTACGCCGCTACCCGATCTCTCC</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>h31.m1R</td>
<td>CGCAAAAGGCAACAAATTGCGAGT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>h31.m2F</td>
<td>CGAAGGCGATCAGATGTGCAAG</td>
<td>815</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td>h31.m2R</td>
<td>ACTGTGCGAGACGTGGAAGG</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. The primers used for multilocus sequence typing to characterize *C. flavescens* strains, all designed in this study except ITS5 and ITS4 (White 1990).
<table>
<thead>
<tr>
<th>Scaffold accession No.</th>
<th>Augustus prediction</th>
<th>JGI prediction</th>
<th>Blastx top hits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length (bp)</td>
<td>Start-stop</td>
<td>Exon/intro number</td>
</tr>
<tr>
<td>CAUG 01000126</td>
<td>4753</td>
<td>34,876-39,628</td>
<td>7/6</td>
</tr>
<tr>
<td>CAUG 01000141</td>
<td>4145</td>
<td>34,763-38,907</td>
<td>2/1</td>
</tr>
<tr>
<td>CAUG 01000321</td>
<td>3852</td>
<td>27,901-31,752</td>
<td>1/0</td>
</tr>
<tr>
<td>CAUG 01000531</td>
<td>3470</td>
<td>192,870-189,401</td>
<td>8/7</td>
</tr>
</tbody>
</table>

Table 6. Prediction of putative non-ribosomal peptide synthetase genes by Augustus, JGI and Blastx.
<table>
<thead>
<tr>
<th>Scaffold accession No.</th>
<th>Augustus prediction</th>
<th>JGI prediction</th>
<th>Blastx top hits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Length (bp)</td>
<td>Start-stop</td>
<td>Query coverage</td>
</tr>
<tr>
<td>CAUG 01000244</td>
<td>2735</td>
<td>75,588-78,322</td>
<td>44% Cellulase (glycosyl hydrolase family 5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>86-95% Cellulase like glycosyl hydrolase or cytoplasm protein from Cryptococcus as best hits, max identity 66-68%; rest of hits mostly glycosyl hydrolase from various fungi</td>
</tr>
<tr>
<td>CAUG 01000336</td>
<td>2773</td>
<td>28,079-25,307</td>
<td>9/8 56% Glycosyl hydrolase family 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>49-61% Endoglucanase E-4 from Cryptococcus as best hits, max identity 66-68%; rest of hits mostly glycosyl hydrolase from various fungi</td>
</tr>
<tr>
<td>CAUG 01000468</td>
<td>2710</td>
<td>65,926-63,217</td>
<td>3/2 64% Endoglucanase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>60-85% Exo-beta-1,3-glucanase from Cryptococcus as best hits, max identity 68-70%; rest of hits mostly exo-beta-1,3-glucanase or glycosyl hydrolase from various fungi</td>
</tr>
<tr>
<td>CAUG 01000595</td>
<td>3353</td>
<td>11,308-7,956</td>
<td>10/9 40% Beta-glucanase/Beta-glucan synthetase; Beta-glucan synthesis-associated protein (SKN1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>44-56% Glucosidase or hypothetical protein from Cryptococcus as best hits, max identity 61-65%; rest of hits mostly glucosidase or glycosyl hydrolase from various fungi</td>
</tr>
</tbody>
</table>

Table 7. Prediction of putative glucanase genes by Augustus, JGI and Blastx.
Figure 1. Venn diagram showing the numbers of *C. flavescens* 3C scaffolds where features were predicted by Augustus and/or JGI and/or Blastx hits were found.

There were a total of 712 scaffolds in the genome assembly used.
Chapter 3: Monitoring field dispersal and persistence of *C. flavescens* OH182.9_3C using quantitative PCR

**Abstract**

*Cryptococcus flavescens* strain OH182.9_3C (3C) previously exhibited significant biological control efficacy against Fusarium head blight, a globally important disease of wheat. In this study, a quantitative PCR (qPCR) assay targeting one copy of the *hsp70* gene was developed and applied to monitor the population dynamics of 3C-like *C. flavescens* in wheat fields and on harvested grains. The assay is specific to strain 3C, though detection of a different subspecies of *C. flavescens* can occur with more than 10,000-fold lower sensitivity. The utility of the qPCR assay to track the biocontrol inoculant was demonstrated in a field experiment. Strain 3C colonized and persisted on wheat at the magnitude of $10^{4.5}$ to $10^{6.5}$ target gene copies per gram head, a level comparable to the inoculation rate. Furthermore, the applied population apparently dispersed from the inoculated plots to non-inoculated areas in a stochastic fashion during the growing season. Populations of 3C-like *C. flavescens* were detected at levels of $10^{2.5}$ to $10^{5.5}$ copies per g on postharvest residues in inoculated areas during the winter and spring following the growing season, and higher populations were detected on non-rotting wheat stalks than on other sampled components of the field. On harvested grain from inoculated plots, 3C-like *C. flavescens* populations declined postharvest, but persisted for up to 155 d of storage at detectable levels. In contrast, 3C-like populations were found to be approximately 100-fold less on uninoculated wheat at all tested pre- and post-harvest time points. The assay presented can be used to determine the environmental
fate of 3C-like *C. flavescens* as an active ingredient of biopesticide and to study the ecological behavior of 3C-like populations in agroecosystems.

**Introduction**

Fusarium head blight (FHB) is a devastating disease of wheat, corn and barley in humid and semi-humid regions throughout the world. It causes hundreds of millions to billions of dollar losses per year during FHB epidemics in United States (McMullen et al., 1997; Nganje et al., 2001; Nganje & Johnson, 2003). The predominant causal agent of this disease in North America is *Fusarium graminearum* Schwabe (teleomorph: *Gibberella zeae* (Schwein.) Petch) (Aoki & O'Donnell, 1999). This fungal pathogen displays a regional phylogenetic diversity (Akinsanmi et al., 2006; Walker, Leath, Hagler, & Murphy, 2001). It infects wheat heads during anthesis through the soft dough stage of kernel development and causes substantial yield loss as well as contamination of grains by mycotoxins such as deoxynivalenol and zearalenone.

Current strategies of FHB control include the use of chemical fungicides, cultural practices and host resistance. Several classes of chemical fungicide and the practice of conventional tillage were shown to have certain degrees of efficacy against FHB (Beyer et al., 2006; Lori et al., 2009; Paul et al., 2008). However, substantial and consistent control of FHB has not been achieved. Moreover, the residues and high cost of chemical fungicides and the desires for reduced or conservation tillage to prevent soil erosion and runoff (Kushwaha et al., 2001; Leys et al., 2010) discourage the use of those two strategies. Various mechanisms of host resistance have been identified in wheat (Lori et al., 2009; Makandar et al., 2012; Skinnes et al., 2010), but no commercial cultivar with satisfactory resistance to FHB is currently available. Therefore, development of microorganisms as biological control agents against FHB will greatly contribute to the management of this disease.

Previous work has established that the yeast *Cryptococcus flavescens* strain OH182.9, when sprayed during the anthesis of wheat, can significantly reduce FHB damage under a
wide variety of conditions. OH182.9 was able to reduce FHB disease severity by as much as 40% to 50% in green house and 60% to 70% in field (Boehm et al., 1999; Khan et al., 1999a; Khan et al., 1999b; Khan et al., 2001; Khan et al., 2004). It also reduced disease incidence and deoxynivalenol accumulation in harvested grains and increased kernel yield in both green house and field (Khan et al., 2001; Khan et al., 2004). The aforementioned biocontrol efficacy of OH182.9 varied depending on factors in experimental design such as wheat cultivar, pathogen isolate, conditions of inoculum preparation, timing of inoculum application and field locations (Khan et al., 2001; Khan et al., 2004; Schisler et al., 2000; Schisler et al., 2002). OH182.9_3C (3C) is a fungicide-resistant variant of OH182.9 and demonstrated higher level of biocontrol efficacy than OH182.9, therefore 3C can be applied in combination with fungicide to improve disease control (Schisler et al., 2009). 3C is able to effectively colonize wheat heads (Schisler et al., 2009; Schisler et al., 2010), where the *F. graminearum* infect and cause symptoms. Currently, a biopesticide with 3C as the active ingredient is being developed through a partnership between The Ohio State University, USDA-ARS, and a private company.

As part of the registration process of 3C as a biopesticide with the U.S. Environmental Protection Agency (EPA), environmental fate determination as one aspect of ecological risk assessment of 3C needs to be conducted (Data Requirements for Pesticide Registration, EPA [http://www.epa.gov/pesticides/regulating/data_requirements.htm]). Therefore, an assay is needed to monitor the 3C dispersal and persistence in the field. Quantitative real-time polymerase chain reaction (qPCR) is a molecular method for fast and sensitive detection as well as precise quantification of target nucleic acid sequences. It has been widely used for quantification of microbial population in the environment. A qPCR assay targeting a conserved region in 3C genome will be suitable for the monitoring purpose in this study. Several other *C. flavescens* strains showed 99.8% to 100% sequence identity to 3C at multiple putative conserved regions including ITS1-5.8S-ITS2, β-tubulin, chitin synthase, elongation factor 1 and heat shock protein 70 kDa (X. Rong, D. A. Schisler and B. B. McSpadden Gardener, unpublished data). Therefore, the highest achievable specificity of
qPCR assays targeting such conserved regions is expected to 3C-like *C. flavescens* instead of to 3C only.

Dispersal of biocontrol agents on host plants and dispersal of fungal pathogens on wheat have been studied in different plant-microbe systems by artificially inoculating the microorganisms in certain areas of field (Cowger, Wallace, & Mundt, 2005; Frantzen & van den Bosch, 2000; Johnson, Stockwell, Sawyer, & Sugar, 2000; Upper, Hirano, Dodd, & Clayton, 2003). The centering of inoculation areas (source) surrounded by non-inoculation areas (sink) enabled the monitoring of dispersal along various distances and to different directions from the sources (Cowger et al., 2005; Johnson et al., 2000). Statistical power of analysis was enhanced by replication of source-sink units across the fields (Cowger et al., 2005; Frantzen & van den Bosch, 2000; Johnson et al., 2000). Therefore, an experimental design of three replicated source areas surrounded by sink areas was used in our study.

The objectives of this study were to (i) develop a qPCR assay of SYBR® Green chemistry to quantify populations of 3C-like *C. flavescens* on field-grown wheat samples and determine the suitability of this assay to (ii) investigate the temporal and spatial population dynamics of 3C-like *C. flavescens* in a wheat field inoculated with 3C, and (iii) examine the persistence of 3C-like *C. flavescens* on harvested grains and postharvest field residues. The data of objectives (ii) and (iii) can be used to assess whether the assay developed in objective (i) is able to detect meaningful patterns of population dynamics. If yes, the assay can be used in future determination of environmental fate of 3C-like *C. flavescens* populations, and the population dynamics data can provide insight into the ecological characteristics of 3C-like *C. flavescens* in wheat production system.

**Materials and methods**

**Strains**

*Cryptococcus flavescens* strain OH182.9_3C (3C) was obtained from the Northern Regional Research Laboratory (NRRL) under accession number NRRL Y-50378. Six
other *C. flavescens* strains were also obtained from NRRL under accession numbers Y-7372, Y-7373, YB-328, YB-601, YB-602 and YB-744, respectively. All the strains were maintained as glycerol stocks at -80°C for long term storage.

**Field establishment and 3C application**

The experiments of this study were performed in two fields planted with Soft red winter wheat (*Triticum aestivum* L.) in the Snyder Farm at the northern end of Wooster, Ohio. The first field, subsequently noted as Field One, was planted with Cultivar Hopewell on October 11, 2010. The layout of Field One is shown in Figure 2. The second field, subsequently noted as Field Two, was planted with Cultivar Freedom on October 6, 2011. Wheat in Field Two was planted in a rectangle area of 85 m × 16 m containing 14 × 7 plots of 4.6 m × 1.7 m per plot with spaces of 1.5 m along the width and 0.6 m along the length. Corn kernels colonized by *Fusarium graminearum* were distributed evenly throughout Field Two approximately three weeks prior to anthesis. The data shown in Figure 4, Figure 6 and Figure 7 were collected from Field One and Two, respectively.

**Inoculum production**

The method of producing 3C inoculum for the Field One was similar to that in (Dunlap & Schisler, 2010) and (Schisler, Slininger, Boehm, & Paul, 2011). Briefly, log-phase cells of 3C produced in SDCL medium (Slininger et al., 2010) served as a 5% seed inoculum for the production in B Braun D-100 fermentor charged with 80 liters of SDCL medium. The fermentor was operated at 25°C, 20 liter/min aeration and 200 revolution/min with agitation provided by twin Rushton impellers. Eighteen hours after inoculation, fermentor temperature was reduced to 15°C to cold shock the cells for 24 h prior to harvest. The cells were harvested through concentration into a paste using a Sharples 12-V tubular bowl centrifuge. The cell paste was then suspended in weak PO₄ buffer (Schisler et al., 2002) at approximately 2 × 10⁹ CFU/ml and frozen at -20°C as stock. The field inoculum was prepared right before application by mixing 1.5 liter of stock (pre-thawed at 10°C overnight) with 1.5 liter sterile double distilled water and 10.8
ml 10% Tween 80. During anthesis of the wheat (Feekes Stage 10.5, on May 29, 2011), three areas of 6.1 m by 6.1 m (Plots i_1, i_2, and i_3 of Figure 2) were sprayed with the 3C inoculum of approximately $1.5 \times 10^9$ cells/ml at the rate of $2 \times 10^6$ to $6 \times 10^6$ cells per square centimeter.

The method of producing 3C inoculum for Field Two was similar to that in (Schisler et al., 2011). Briefly, 500-ml flasks containing 200 ml SDCL medium inoculated with 3C (starting OD of 0.10 at $A_{620}$) were incubated at 25ºC at 250 revolution/min for 24 h. These cultures were then used to seed 3-liter Fernbach flasks containing 1.25 liter of SDCL medium to an OD of 0.10. Cultures were incubated for 48 h at 25ºC and 250 revolution/min. Colonized broth of $2.3 \times 10^8$ CFU/ml was then transferred into sterile containers, transported to the field on ice and applied at the rate of $4.4 \times 10^5$ CFU per square centimeter in the field within 24 h. The application was performed during anthesis of the wheat (Feekes Stage 10.5, on May 29, 2012). Sampling for this study was performed in 12 plots, 6 of which were sprayed with 3C only and another 6 of which were not sprayed with any microorganism. Some other plots in the field were also sprayed with 3C but they were all sprayed with other microorganism as well. The plot locations of different treatments were randomized across the field.

**Sample collection**

Wheat head samples were collected by clipping at the stalk 1 to 2 cm below the head and sealing each head in an individual sample bag. During sample collection, each head was only touched with the inside of the bag used to collect it and the clipping scissors were sterilized with 75% aqueous ethanol between clips to prevent cross-contamination. Heads were collected from Field One immediately after the spray application of 3C (0 day post inoculation) as well as 24 h (1 day), 10, 26 and 44 days post inoculation in the summer of 2011 (see Figure 2 for details on the locations of sampling). Heads were also collected from Field Two at 11 days and 1 day prior to harvest on June 26, 2012, where one head was collected from each of the 12 plots at each time point.
Postharvest residue samples were collected from the three inoculated areas of Field One in late January and late March of 2012. Seven different locations were sampled within a circle of about 1-m radius around the center of each inoculated area. At each time point, two types of samples were collected from each location: a mixture of soil, living weeds and brownish rotting plant materials was collected as one composite sample, and non-rotting wheat stalks were collected as the other sample.

The grains were harvested from each plot as a whole for the 12 plots of Field Two on June 26, 2012. The total grains of each plot were stored in one paper bag as a layer of thinner than 6 cm at dark at room temperature. One subsample of approximately 10 g was collected from the total grains of each plot immediately prior to grinding up for DNA extraction at 14, 64 and 155 days postharvest, respectively.

Sample handling, DNA extraction and template preparation

Genomic DNA of 3C and the six other *C. flavescens* strains was extracted from 2-day old cultures grown on 1/5 Tryptic Soy Agar using PowerSoil™ DNA Isolation Kit (MO BIO Laboratories, Inc., California, USA) following the manufacturer’s instruction. These original DNA extracts were diluted by more than 20-fold to certain concentrations based on the measurements on NanoDrop ND-1000 (Thermo Scientific, Delaware, USA) to prepare the templates for qPCR.

The wheat head and stalk samples were returned to the lab within 1 hr of sampling and stored at -80°C. The wheat (including grain) samples were ground up in liquid nitrogen immediately prior to DNA extraction. The composite residue samples were stored at 10°C after brought back to the lab and then DNA extraction was done within 24 h directly from the samples without grinding in liquid nitrogen. Total DNA was extracted from approximately 0.3 gram of each sample using PowerSoil™ DNA Isolation Kit following the manufacturer’s instruction, yielding 100 µl DNA extract per sample. Since the DNA concentrations of the wheat head and grain samples were so high (due to the abundant wheat DNA from fresh tissue) as to interfere with the interpretation of qPCR data, 1:20 dilutions and 1:10 dilutions of the original DNA extracts were used as
templates for those samples from Field One and Two, respectively. Non-diluted DNA extracts were used as the templates of postharvest residue samples from Field One.

**Quantitative PCR (qPCR)**

qPCR primers were designed using Primer3 (Rozen & Skaletsky, 2000) based on the draft genomic sequences of 3C (EMBL accession number CAUG01000001 to CAUG01000712). qPCR test for specificity was performed on 14 pairs of primers targeting 7 putative conserved regions, using the 1 ng/µl genomic DNA of 3C and the other 6 *C. flavescens* strains as template. The 7 regions were putatively ITS1-5.8S-ITS2 and partial regions of the following protein encoding genes: β-tubulin, chitin synthase 5, elongation factor 1 and three heat shock protein 70 kDa. Among the 14 pairs of primers, Pair h31.2 showed the greatest difference in Threshold Cycle values between 3C and any of Y-7372, YB-328 and YB-744, and was thus chosen for use in the qPCR assay described as follows. Each qPCR reaction had a total volume of 25 µl, consisting of 2.5 µl template DNA, 0.4 pmol/µl of each primer (forward primer h31.2_F: 5'-CGTCAGCGTGTTGCCACTTCGT-3'; reverse primer h31.2R: 5'-GCTGCTGTCTTGCGGTCGCTA-3’), 12.5 µl iQ™ SYBR® Green Supermix (Bio-Rad Laboratories, Inc., California, USA) and DNA-free PCR water as the rest of the volume. DNA-free PCR water was also used as the template of negative controls. The thermal cycling program was run on the iCycler and the fluorescence data was collected by the iQ™5 detection system (Bio-Rad Laboratories, Inc.). Amplification was two-step with 3 min at 95°C followed by 40 cycles of 15 s at 95°C and 1 min at 72°C. Then the melting curve was run from 55°C to 95°C with an increment of 0.5°C and a dwelling time of 10 s at each temperature. The fluorescence data of qPCR was analyzed using the iQ™5 Optical System Software (Bio-Rad Laboratories, Inc.).

**Data processing and analysis**

The standard series of 1, 1 × 10⁻¹, 1 × 10⁻²…1 × 10⁻⁷ ng/µl 3C genomic DNA in template were run in triplicate to obtain the regression equation \( Y = 5.792 - 0.2984X \) \((R^2 = 0.991)\). This equation was used to calculate from the field samples’ Threshold Cycle
values \((X)\) their log starting quantity values \((Y)\) as equivalences to 3C genomic DNA concentration in template. The unit of starting quantity was converted from “ng/µl 3C genomic DNA in template” to “target gene copy number per gram sample”. This unit conversion was based on the following two assumptions: (i) there was one copy of target gene per genome, supported by BLAST search of this gene and the h31.2 primers against the draft genome of 3C and by the sequencing coverage of the respective scaffold relative to the coverage of the whole genome (data not shown); (ii) the genome size of 3C was about 0.02 pg per haploid genome (estimated based on the genome sizes of Cryptococcus gattii and C. neoformans) (B. Kullman, H. Tamm and K. Kullman, 2005, Fungal Genome Size Database [http://www.zbi.ee/fungal-genomesize]). Accordingly, 1 ng 3C genomic DNA was considered equal to \(5 \times 10^4\) copies of target gene and, reciprocally, one copy of target gene equal to \(2 \times 10^{-5}\) ng 3C genomic DNA. In this study, each reaction started with 2.5 µl template, which belonged to a total of 100 µl DNA extract resulted from 0.3 gram sample. Therefore, 1 ng/µl 3C genomic DNA was considered equal to \(3.3 \times 10^8\) \((10^{8.5})\) or \(1.7 \times 10^8\) \((10^{8.2})\) copies of target gene per gram sample for the wheat head or grain samples (1:20 or 1:10 dilutions of original DNA extracts as templates, respectively) and \(1.7 \times 10^7\) \((10^{7.2})\) copies per gram sample for postharvest residue samples from Field One (non-diluted DNA extracts as templates).

The theoretical limit of detection for any given PCR reaction was considered to be one copy of target gene per reaction. Given the two aforementioned assumptions for unit conversion and that each reaction started with 2.5 µl template in this study, the theoretical detection limit of this study was \(2 \times 10^{-5}\) ng 3C genomic DNA per 2.5 µl, equal to \(8 \times 10^{-6}\) ng/µl 3C genomic DNA in template. For the field samples, of which a total of 100 µl DNA extract resulted from 0.3 gram sample, the theoretical detection limits were \(2.7 \times 10^3\) \((10^{3.4})\) or \(1.3 \times 10^3\) \((10^{3.1})\) copies of target gene per gram sample for the wheat head or grain samples (1:20 or 1:10 dilutions of original DNA extracts as templates, respectively) and \(1.3 \times 10^2\) \((10^{2.1})\) copies per gram sample for the postharvest residue samples from Field One.

In the melting curve analysis for qPCR, a single peak of 87.0 or 87.5°C was the only peak reported by the iQ™ detection system for the concentrated templates in the
standard series (1 through $10^{-4}$ ng/µl 3C genomic DNA). Therefore for field samples, only those where a single peak of 87.0 or 87.5°C was the only reported peak were considered as positive for detection of 3C. The samples showing no peak of 87.0 or 87.5°C were considered as no detection of 3C and their Ct values were considered as 41 for statistical analysis. The samples showing a peak of 87.0 or 87.5°C as well as other peaks were trimmed out of further analysis since they showed detection of 3C but the quantification will not be precise enough due to the interference of non-specific peaks. A few statistical outliers were also trimmed out. For each experiment, the trimmed-out data points never exceeded 7% of the total data points, unless specified. Analysis of variance using General Linear Model where $\alpha = 0.05$ was performed on the ranked forms of Ct values using the software Minitab 16 (Minitab Inc., Pennsylvania, USA).

Results

Sensitivity and specificity of the assay

The primer pair h31.2 used in the qPCR assay were designed to amplify part of a putative heat shock protein 70 kDa ($hsp70$) gene identified from the draft genomic sequence of 3C. Primer h31.2F binds to a putative intron region and primer h31.2R binds to a putative exon region (Figure 3). The target region was 150 bp and includes one putative exon-intron conjunction. The amplification efficiency of the qPCR assay ranged from 98% to 109%. The threshold cycle (Ct) values and melting temperatures of selected qPCR reactions were listed in Table 8. The standard series displayed a gradient of increasing Ct values from template concentration 1 to $1 \times 10^{-4}$ ng/µl, while there was no detection for $1 \times 10^{-6}$ or $1 \times 10^{-7}$ ng/µl. Among the three replicates of $1 \times 10^{-5}$ ng/µl, one showed detection while the other two did not. Since the theoretical detection limit was $0.8 \times 10^{-5}$ ng/µl 3C genomic DNA in the template (equivalent to one copy of target gene per reaction), it is reasonable that there was no consistent detection when the template concentration of 3C genomic DNA was $1 \times 10^{-5}$ ng/µl or less. Since the error of pipetted amount of target DNA is very high when the target concentration is near one copy per
reaction, any detection event by Ct greater than 35.1 (the Ct of that standard rep at $1 \times 10^{-5}$ ng/µl) should be considered only qualitative instead of quantitative.

To test the specificity of this assay within the Cryptococcus flavescens species, qPCR was also run on the 1 ng/µl genomic DNA templates of six other C. flavescens strains than 3C (Table 8). The sequences of strains Y-7372, YB-328 and YB-744 differ from 3C by two base pairs at each primer binding site (Figure 3). These three strains showed much higher Ct values than 3C genomic DNA at the same template concentration. The Ct difference of around 15 (equal to approximately $3 \times 10^4$-fold difference in starting quantity) between 3C and the aforementioned three strains can well differentiate 3C from them. In contrast, strains Y-7373, YB-601 and YB-602 showed 100% sequence identity to 3C across the whole amplicon (Figure 3). These three strains could not be differentiated from 3C by the assay primer pair h31.2 (Table 8) or any of the other 13 tested primer pairs (data not shown) since they showed comparable Ct values to 3C for all 14 primer pairs. Therefore, what the assay quantifies is the population of 3C-like C. flavescens, including 3C as well as Y-7373, YB-601 and YB-602, instead of 3C population only.

**Dynamics of 3C-like C. flavescens population in areas inoculated with 3C**

The population density of 3C-like C. flavescens on wheat heads, represented by the target gene copy number per gram wheat head tissue, fluctuated over time in the areas inoculated with 3C (Figure 4). The samples at one day post inoculation (DPI) showed significantly lower population than those at and after 10 DPI. There was an obvious trend that population decreased within the day immediately after inoculation and then increased during between 1 and 10 DPI. This indicates the DNA of 3C applied to the wheat heads was partially degraded after inoculation but the surviving 3C cells also started to multiply on wheat heads, resulting in the qPCR detection of population growth later on. There was a slight trend of progressive population increase in from 10 to 26 and to 44 DPI but the increase was smaller than that from 1 to 10 DPI. This indicates that later in the season the population of 3C-like C. flavescens was still growing but the growth rate was reduced.
This was probably due to the saturation of wheat head tissue as colonization niche for 3C-like *C. flavescens* later in the season.

3C-like *C. flavescens* population on the majority of the samples was between $10^{5.5}$ and $10^{6.0}$ copies of target gene per gram wheat head tissue immediately post inoculation (0 DPI). Subsequently, population decreased to between $10^{4.5}$ and $10^{5.5}$ 1 d DPI and then gradually increased to between $10^{6.0}$ and $10^{6.5}$ by the end of the season. Given that the inoculation rate was $10^{6.3}$ to $10^{6.8}$ 3C cells per square centimeter and that the horizontal section of a wheat head was around one square centimeter, qPCR-detected population was maintained at a level that was slightly lower than, but still comparable to, the inoculation rate throughout the growing season.

**Dispersal of 3C to the non-inoculated area of field**

There was a trend that 3C-like population decreased as the distance from inoculated area increased and population in non-inoculated area increased over time from 1 to 44 DPI (Figure 5A and Figure 5B). This indicates that 3C was dispersed stochastically from the inoculated area to non-inoculated area over time. By the last sampling in the growing season (44 DPI), the majority of samples from non-inoculated area harbored $3.3 \times 10^3$ to $3.3 \times 10^5$ target gene copies per gram wheat head of the *hsp70* marker.

There were a few samples (i.e. 5 of 36) positive for 3C-like *C. flavescens* in the non-inoculated area at 0 DPI. However the level detected was $3.3 \times 10^2$ to $3.3 \times 10^3$ target gene copies of the target per gram wheat head, an amount around the detection limit of the assay. And, none of the non-inoculated samples at 1 DPI (i.e. 0 of 35) scored positive for 3C-like populations. These data indicate that colonization of wheat heads by native *C. flavescens* was generally infrequent and, when it did occur, never reached $10^4$ gene copies per wheat head. Thus, the majority of qPCR-detected populations in this experiment were interpreted as arising from 3C inoculation, and this is why the aforementioned population dynamics over distances and time were believed to reflect the dispersal pattern of 3C.

**The effect of sample collection activity on 3C dispersal**

50
To test whether sample collectors walking throughout the field contributed to the apparent dispersal of 3C, additional wheat head samples were collected from borders of wheat-grown area in Field One on 10 DPI. Since sample collectors walked along the fallow walkways in between the wheat-grown area, it was possible that they brought 3C from inoculated area to non-inoculated area. It was hypothesized that if sample collection activity did contribute to 3C dispersal, the wheat heads at the borders of wheat-grown area neighboring the walkways should show higher 3C population density than those deeper inside of wheat-grown area do. The samples relevant to this hypothesis were collected from the eastern side of the field (see Figure 2 for details on sampling locations) since the sample collectors were required to exit the field from that side after they collected samples from inoculated area. There was no significant difference ($n > 7, P > 0.78$) between the population densities on the wheat heads from the borders and inside of wheat-grown area, indicating that the sample collection activity did not significantly contribute to the dispersal of 3C in this experiment.

**The effect of prevailing wind direction on 3C dispersal**

To test whether the direction of prevailing wind (from west to east) affected the apparent patterns of 3C dispersal, the wheat head samples from the western and eastern sides of Field One were compared. The samples from different location groups (see Figure 2 for details on location groups) and/or at different time points were compared separately. It was hypothesized that the samples from the eastern side of the field should have shown higher 3C population density than those from the western side did if the prevailing wind did play an important role in 3C dispersal. Actually the differential patterns of 3C population density were not consistent across different location groups and sampling time points, and there was no obvious trend towards either side of the field (data not shown). This indicates that the direction of prevailing wind did not have a significant effect on the patterns of 3C dispersal in this experiment. This is consistent with the fact that 3C is a yeast without known production of wind-dispersed spores,
therefore may not be easily dispersed by wind alone. Though 3C may be better dispersed by wind-blown rain, rainfall events may not correlate with prevailing wind direction.

**Overwintering persistence of 3C-like *C. flavescens***

To test whether 3C-like *C. flavescens* can persist in the field over the winter, the three inoculated areas of Field One were sampled in the winter to early spring of 2012. The data showed that population could still be detected in inoculated area by 6 to 9 months postharvest (Figure 6). Since there was no significant difference among the same type of samples in different inoculated areas (data not shown), the data points from the three inoculated areas were combined for the comparison between sample types. The non-rotting wheat stalk samples showed significantly higher levels of population (mostly $10^4$ to $10^5$) per gram weight than the composite samples of soil, living weeds and rotting plant materials (mostly $10^2.5$ to $10^3.5$) did (Figure 6). Population on overwintering residues was lower than that on wheat heads by the end of growing season, which was mostly above $10^6$ target gene copies per gram sample at 44 DPI.

**Population levels of 3C-like *C. flavescens* around harvest and during storage**

Additional wheat head and grain samples were collected from another field (Field Two) at different time points around harvest. Population levels on grains were repetitively measured by qPCR after periods of storage. The inoculated samples showed generally higher population than the non-inoculated samples for both sample types at different time points (Figure 7). Most of the samples from inoculated plots, including the heads and the threshed grains, showed detection of higher than $10^{3.5}$ and up to $10^{6.7}$ target gene copies per gram. There was a trend that population on wheat heads increased from 11 days pre-harvest to 1 day pre-harvest in both inoculated and non-inoculated plots. For the samples collected from non-inoculated plots, population was significantly lower on the threshed grains of 14 days postharvest (mostly around $10^{3.5}$ target gene copies per gram) than on the heads of one day pre-harvest (mostly between $10^{4.0}$ and $10^{4.5}$ target gene copies per gram) ($P = 0.037$ for Tukey’s test). The inoculated samples showed a trend of such difference but it wasn’t significant ($P = 0.22$ for Tukey’s test). There was a
trend of progressive population decrease on threshed grains from both inoculated and non-inoculated plots during storage from 14 to 64 and to 155 days postharvest (Figure 7).

**Discussion**

**Effectiveness of the qPCR assay**

Here we developed a qPCR assay to monitor the dispersal and persistence of 3C-like *Cryptococcus flavescens* on field-grown wheat and, more broadly, in the environment. This assay enabled the detection with clear differential between the inoculated and non-inoculated wheat head samples (Table 8, last two rows). And it can be reasonably applied to multiple sample types, since it gave meaningful differential results within each of the sample types tested in this study. The sensitivity of our assay was comparable to other related qPCR assays. Previous studies using qPCR to detect fungal pathogens on wheat showed minimum detections at the magnitude of $10^1$ gene copies per reaction and $10^4$ gene copies or 1 to $10^3$ ng genomic DNA per gram wheat tissue (Hogg, Johnston, & Dyer, 2007; Hogg et al., 2010; Knight, Sutherland, Martin, & Herde, 2012; Palazzini, Groenenboom-de Haas, Torres, Köhl, & Chulze, 2012), which were comparable to those in our assay (12.5 gene copies per reaction and $10^3$ to $10^4$ gene copies or 0.01 to 0.1 ng genomic DNA per gram wheat tissue) (Table 8, Figure 5B and Figure 6). Our qPCR assay may not be as sensitive as culture-based quantification of 3C, which could detect 10 to 100 cells per gram wheat tissue (Schisler et al., 2009; Schisler et al., 2010). However, this qPCR assay is still superior to culture-based quantification in that the former has inter-strain specificity while the latter cannot differentiate among subspecies of *C. flavescens* based on morphology. Also the qPCR assay enables long-term storage of samples prior to quantification without underestimating the populations due to die-off of cells during storage.

The qPCR assay showed intra-species specificity to four 3C-like *C. flavescens* strains (3C, Y-7373, YB-601 and YB-602) over the other three strains (Table 8). No further discrimination by Threshold Cycle (Ct) value nor melting temperature can be achieved.
among 3C-like strains, which is due to the high sequence similarity among them. Sequence difference between 3C and the three 3C-like strains was actually 0 to 1 bp across a total of more than 5500 bp of DNA (including ~2000 bp non-coding sequences) across 9 conserved regions, including all the 7 regions used for primer design in this study (X. Rong, D. A. Schisler and B. B. McSpadden Gardener, unpublished). Since mismatch by one base pair at primer binding site only gave rise to a Ct difference of ~1.5 for one of the tested primers (data not shown), it is almost impossible to differentiate among 3C-like C. flavescens through amplification from those conserved regions. It may be possible to design primers targeting more variable regions to achieve “3C-only” specificity. However, such a strategy may sacrifice sensitivity because the higher natural mutation rates in variable regions may result in loss of ability to detect 3C progenies that have mutations in primer binding sites. Furthermore, the possibility cannot be excluded that some 3C-like strains may have originated from the same population as 3C in the near past and were named as different strains simply due to separate isolation events. Recent work in our labs indicates that the biocontrol efficacy against Fusarium head blight did not differ significantly among 3C and other 3C-like strains (X. Rong, D. A. Schisler and B. B. McSpadden Gardener, unpublished). Therefore, exhaustive pursuit of additional “3C-only” specificity will not provide much scientific or practical value.

Whether the detected 3C population levels by our assay are well correlated with the viability of 3C cells needs further investigation. In the Reference (Schisler et al., 2009) and (Schisler et al., 2010), applying 3C inoculum of approximately $3 \times 10^8$ CFU/ml at the rate of about $10^6$ CFU per square centimeter resulted in $10^1$ to $10^2$ CFU/g glume tissue at 16 h post inoculation and $10^5$ to $10^6$ CFU at 256 to 280 h post inoculation. In those studies, strain CFU of strain 3C decreased by a magnitude of $10^4$ to $10^5$ compared to the application rate in the 16 h post inoculation. In contrast, there was only an approximately 10-fold decrease in 3C DNA population during the one day following inoculation in our study (Figure 4). This dramatic difference is possibly due to that most of the 3C DNA detected on 1 DPI in our study belonged to non-viable cells or was already released from lysed cells. However, 3C CFU level increased to the similar level to the application rate 10 days later in those previous studies and was also the case in our
study. This indicates that most of the 3C DNA detected on 10 DPI in our study belonged to viable cells. CFU-based assay to test the viability of 3C cells on wheat heads, field residues and threshed grains is currently in process with side by side comparison to qPCR quantification.

Previously developed qPCR assays of other wheat-associated eukaryotes showed inter-species specificities (Hogg et al., 2007; Knight et al., 2012). Our assay showed a certain level of inter-strain specificity within the species of Cryptococcus flavescens. Specifically, assays of six other isolates of C. flavescens indicated that the species encompassed two genotypic groups based on the amplification efficiency of standard amounts of DNA template (Table 8). Even though all 3C-like isolates showed similar amplification efficiency, we believe that detection of native populations of 3C’s close relatives was minimal in this study. This is because no detection events were noted across nearly all the non-inoculated samples on 0 and 1 DPI while the inoculated samples all showed clear detection (Figure 5). If there had been a significant interference, those non-inoculated samples might have shown more frequent and higher levels of detection of amplification. Second, the level of detection generally decreased as the distances from inoculated plots increase (Figure 5). If there had been a significant interference, the levels of detection should have been generally the same or more random across the field. These data indicate that native populations of 3C-like C. flavescens are relatively rare on wheat heads between flowering and harvest, but more quantitative work in this area seems warranted.

**Population dynamics of 3C-like C. flavescens in the field**

It is well known that microbial populations die off when they reach the death phase of growth. In some studies on the population dynamics of plant-associated bacteria using culture-based methods (Johnson et al., 2000; Lindow & Suslow, 2003; Upper et al., 2003), decreased bacterial population on flowers or leaves of host plants later in the experimental time courses were observed. In a study using qPCR for population quantification (Hogg et al., 2010), the total population of three Fusarium species died off significantly on spring wheat stubble residue in six months after the harvest in summer.
In the previous studies on 3C colonization of wheat using CFU-based quantification (Schisler et al., 2009; Schisler et al., 2010), slight die-off of 3C population was observed between 184 and 256 to 280 h after wheat flowering and inoculation in the 2009 run but not in the 2010 run. Our study showed population decrease on wheat heads from 0 to 1 DPI (Figure 4) and on threshed grains during storage (Figure 7). The die-off of 3C-like C. flavescens (the majority of which we believe was 3C) in our study is consistent with the observations in previous studies. This suggests that 3C will probably not pose a threat to ecological balance or food safety through prolonged multiplication.

The dispersal pattern of 3C across the wheat field in this study resembles those reported previously for several fungal pathogens on wheat (Cowger et al., 2005; Frantzen & van den Bosch, 2000). In such studies, at each given time point, the microorganism population decreased steeply from artificially inoculated area (source) to non-inoculated area (sink) near the source, while decreased gently over farther distances across the sink, especially at early time points. In our study, there were obvious increases in 3C population at the same distances in the sink over time (Figure 5), which is more similar to the dispersal pattern of a wheat pathogen on pure stand of susceptible cultivar than to on a mixture of susceptible and resistant cultivars (Cowger et al., 2005). This suggests that wheat is a compatible host to 3C, which is consistent with the fact that 3C is able to colonize wheat heads effectively (Schisler et al., 2009; Schisler et al., 2010). And our assay is robust enough to successfully demonstrate this pattern.

Possible application of the qPCR assay

In previous studies, qPCR assays for quantification of wheat pathogens have been used to evaluate disease severity and to monitor and predict disease development (Hogg et al., 2007; Hogg et al., 2010; Knight et al., 2012). Our qPCR assay can be used to determine environmental fate of 3C-like C. flavescens as biopesticides. Specifically, dispersal and persistence of 3C-like C. flavescens can be investigated throughout the wheat production system at various geographic locations over multiple growing seasons. It can also be used to study the effects of common field abiotic and biotic factors of concern on 3C population dynamics in order to optimize the formulation of 3C-based
biopesticides. Further, it can be used to investigate the association between 3C population dynamics and biocontrol efficacy of 3C against Fusarium head blight under various conditions. This will facilitate the development of 3C as a biopesticide and provide guidance in the frequency and timing of 3C application. And the study on the correlation between the qPCR- and culture-based population quantification of 3C will enable better interpretation of the qPCR assay results. In addition, the target genes of this assay and other conserved regions of multiple *C. flavescens* strains can be sequenced to develop related qPCR assays that have different levels of specificity to serve different research purposes.
Figure 2. Field map and major sampling scheme for the detection of *C. flavescens* strain 3C in Field One.
Wheat (cultivar Hopewell) was planted throughout the field of 42.7 m by 61.0 m (the biggest rectangle), excluding the area of fallow walkways (indicated by double solid lines), which were 0.6-m wide between each 6.1-m stripes of wheat-grown area. Three areas of 6.1 m by 6.1 m were sprayed with 3C during anthesis of the wheat, indicated by the shaded areas labeled as “i_1”, “i_2” and “i_3” on the map. In the summer of 2011, sampling was done within a rectangle of 30.5 m by 42.7 m (consisting of 5 × 7 square grids of solid lines) by clipping off one wheat head from each of the locations indicated approximately by the Roman numerals (inside of wheat-grown areas) and short thick lines (along the borders of wheat-grown areas neighboring walkways) in the map. Each unique Roman numeral indicates one group of sampling locations of the similar distances to the nearest points on inoculated areas. The mean distances of each group were: I – 0 m, II – 1.60 m, III – 4.64 m, IV – 7.38 m, V – 12.37 m, VI – 14.72 m and VII – 18.12 m. Samples were collected from the locations of Group I to V on 0, 1, 10, 26 and 44 day(s) post inoculation. Group V, VI and VII were sampled only at 44 days post inoculation. The border locations (indicated by short thick lines) were only sampled at 10 days post inoculation.
Figure 3. Sequence alignment of the qPCR amplicons from 3C and other six C. flavescens strains.

The consensus sequence is on the top of the alignment. Primer binds h31.2_F (1 to 22 bp) and h31.2_R (129 to 150 bp) as well as the region predicted by Augustus as exon (39 to 150 bp) are indicated.
<table>
<thead>
<tr>
<th>Nature</th>
<th>Strain name or other identifier</th>
<th>Concentration (ng/µl in template)</th>
<th>Threshold Cycle&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Melting temperature (&lt;sup&gt;°&lt;/sup&gt;C)&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genomic DNA, mean of 3 replicates of standard series</td>
<td>3C</td>
<td>1</td>
<td>19.4</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10&lt;sup&gt;-1&lt;/sup&gt;</td>
<td>22.5</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10&lt;sup&gt;-2&lt;/sup&gt;</td>
<td>26.2</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10&lt;sup&gt;-3&lt;/sup&gt;</td>
<td>29.7</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10&lt;sup&gt;-4&lt;/sup&gt;</td>
<td>33.0</td>
<td>87.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10&lt;sup&gt;-5&lt;/sup&gt;</td>
<td>35.1 (1 rep) or n.d. (2 reps)</td>
<td>87.0 (1 rep) or n.a. (2 reps)</td>
</tr>
<tr>
<td></td>
<td>10&lt;sup&gt;-6&lt;/sup&gt;</td>
<td>n.d.</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10&lt;sup&gt;-7&lt;/sup&gt;</td>
<td>n.d.</td>
<td>n.a.</td>
<td></td>
</tr>
<tr>
<td>Genomic DNA</td>
<td>Y-7372</td>
<td>1</td>
<td>34.0</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>Y-7373</td>
<td>1</td>
<td>17.3</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>YB-328</td>
<td>1</td>
<td>34.8</td>
<td>87.0</td>
</tr>
<tr>
<td></td>
<td>YB-601</td>
<td>1</td>
<td>19.0</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>YB-602</td>
<td>1</td>
<td>18.7</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>YB-744</td>
<td>1</td>
<td>35.0</td>
<td>87.0</td>
</tr>
<tr>
<td>DNA extract of wheat head from field&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Inoculated with 3C</td>
<td>-</td>
<td>26.5</td>
<td>87.5</td>
</tr>
<tr>
<td></td>
<td>Not inoculated with 3C</td>
<td>-</td>
<td>n.d.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

<sup>a</sup> Threshold Cycle was not detected (n.d.) by the software when fluorescence signal was below the threshold at the end of 40-cycle amplification.

<sup>b</sup> Melting temperature was not applicable (n.a.) for the samples for which the detection software reported no peak of 87.0 or 87.5°C (the only reported peak for standard series 1 through 10<sup>-4</sup> ng/µl in template).

<sup>c</sup> The templates of the field samples are the 1:20 diluted DNA extracts from a 3C-inoculated wheat head 44 days post inoculation and from any non-inoculated head at a distance of 25.8 m to the nearest inoculated area on the day of inoculation.

Table 8. Sensitivity and specificity of the qPCR assay targeting the putative *hsp70* gene.
Figure 4. Population dynamics of 3C-like *C. flavescens* on inoculated wheat heads.

3C inoculum suspension was sprayed on wheat at a rate of $2 \times 10^6$ to $6 \times 10^6$ cells per square centimeter. The wheat head samples ($n = 5$ to 9) collected from the three inoculated areas (see Figure 2 for details on sampling locations) at five time points post inoculation were compared. The median populations of the groups that do not share the same letter (a, b) were significantly different based on Tukey’s test ($P \leq 0.05$). “DL” on the Y-axis represents the theoretical detection limit for this experiment. The number of sample that was negative for 3C detection at any given time point was indicated in a square box at the bottom of the plot.
Figure 5. Changing populations of 3C-like *C. flavescens* in a wheat field following inoculation in part of the field.
The percentages of samples above a certain population level in the linear scale of distance (A) and as the breakdown of all the ranges of detection (B). The distance groups were noted below the X-axes (see Figure 2 for grouping information) and the actual values of mean distance of each group were indicated in the dash-line box of Panel (A). There were only data of Group I to IV for 0 to 26 day(s) post inoculation (DPI) while additional data for Groups V to VII are provided solely at 44 DPI. The numbers of samples from each of the location groups at each time point were: I – 5 to 9, II – 18 to 20, III – 8 to 10, IV – 4 to 6, V – 16, VI – 8, VII – 4. The theoretical detection limit of this experiment was $2.7 \times 10^3$ target gene copies per gram wheat head.
Figure 6. Overwintering persistence of 3C-like *C. flavescens* in artificially inoculated area.

Postharvest residues of the three inoculated areas were sampled in late January and late March of 2012. Composite samples (C) of soil, living weeds and brownish rotting plant materials and samples of non-rotting wheat stalks (W) were collected within a circle of about 1-m radius around the center of each inoculated area. “DL” on the Y-axis represents the theoretical detection limit for this experiment. The numbers of samples that were negative for detection for each sample type and time point are indicated in square boxes at the bottom of the plot. Statistical analysis was run separately for the samples from January and those from March on the ranked forms of all the data points present in the figure. The median populations of the groups (*n* = 21) that share the same number before the letter but not the same letter were significantly different based on Tukey’s test (*P* ≤ 0.05).
Figure 7. 3C-like *C. flavescens* populations on wheat heads closely before harvest and postharvest threshed grains.

One sample of each type was randomly collected for each of the 6 inoculated and 6 non-inoculated plots at different time points. “DL” on the Y-axis represents the theoretical detection limit for this experiment. The numbers of samples that were negative for detection for each time-treatment combination are indicated in the square boxes at the bottom of plot. Statistical analysis shown in this figure was run separately for different sample types and different time points on the ranked forms of all the data points present in the figure. The medians of the population groups (*n* = 6) that share the same number before the letter but not the same letter were significantly different based on Tukey’s test (*P* ≤ 0.05).
Chapter 4: Multilocus sequence typing and phenotyping of 15 *C. flavescens* strains

**Abstract**

*Cryptococcus flavescens* strain OH182.9_3C (3C) previously displayed significant biological control activity against Fusarium head blight, a globally important disease of wheat. Detailed study on diversity within *C. flavescens* species has not been reported before. Multilocus sequence typing was performed on strain 3C and 12 other *C. flavescens* strains using six apparently unlinked loci: rDNA-ITS, β-tubulin, chitin Synthase 1, elongation factor 1, heat shock protein 70 kDa and a mostly anonymous region. The rDNA-ITS sequences were identical across all the 13 strains. The other five loci consistently revealed two genotypes, a 3C type (6 strains, including 3C and type strain NRRL Y-1401\(^T\)) and a non-3C type (7 strains), with high bootstrap support. This genotypic split was congruent to a mating type split previously demonstrated by mating test. Biochemical assays using Biolog supported the grouping by genotype and demonstrated slow assimilation of several carbon sources by two strains of 3C type. Type strain Y-1401\(^T\) generally grew more poorly than all the other 12 strains during the whole course of Biolog incubation. In a green-house bioassay, the strains of both genotype showed biocontrol efficacy against Fusarium head blight and 3C-type as a whole showed slightly higher efficacy than non-3C type. This study provided insight into the phylogenetic relationships within *C. flavescens* species and resources for developing genetic and phenotypic markers to differentiate different groups of *C. flavescens* strains.
Introduction

Fusarium Head Blight (FHB) is a globally devastating disease of wheat, causing up to billions of dollar losses in each epidemic year in United States (McMullen et al., 1997; Nganje et al., 2001; Nganje & Johnson, 2003). This disease is predominantly caused by 
Fusarium graminearum
Schwabe (teleomorph: Gibberella zeae (Schwein.) Petch) in North America (Aoki & O'Donnell, 1999). Infection on wheat heads happens mainly during anthesis, resulting in substantial yield loss and mycotoxin contamination of grains.

FHB is currently controlled by chemical fungicides, cultural practices and host resistance in wheat production, however, substantial and consistent control of FHB has not been achieved (Beyer et al., 2006; Kushwaha et al., 2001; Leys et al., 2010; Lori et al., 2009; Makandar et al., 2012; Paul et al., 2007; Skinnes et al., 2010). Therefore, development of biological control strategies using microorganisms against FHB is essential to further improvement of disease management.

Applications of the yeast 
Cryptococcus flavescens
strain OH182.9 (NRRL Y-30216) during the anthesis of wheat was previously shown to significantly reduce FHB damage under various conditions. OH182.9 displayed high levels of colonization on wheat heads (Schisler et al., 2009; Schisler et al., 2010), where the majority of FHB infection and symptoms occur. Application of 3C resulted in decrease of disease severity by up to 50% in green house and 70% in field (Boehm et al., 1999; Khan et al., 1999a; Khan et al., 1999b; Khan et al., 2001; Khan et al., 2004). OH182.9 also reduced disease incidence and mycotoxin accumulation in harvested grains while increased weight per 100 kernels in both green house and field (Khan et al., 2001; Khan et al., 2004). OH182.9_3C (3C) is a fungicide-resistant variant of OH182.9 selected for co-application with fungicide to enhance control of FHB (Schisler et al., 2009). 3C actually demonstrated improved efficacy against FHB compared to its progenitor OH182.9 (Schisler et al., 2009).

Characterization of the diversity within groups of biocontrol agents enables bioprospecting and biogeographic studies of those organisms. Such work has been done with bacteria (McSpadden Gardener et al., 2000) but to date applications to eukaryote biocontrol agents has been lacking. Cryptococcus flavescens has been isolated from
various substrates including the atmosphere, corn kernels and wheat anthers (Khan et al., 2001; Kurtzman, 1973; Saito, 1922). Two inter-compatible and self-incompatible mating types have been identified in *C. flavescens* (Kurtzman, 1973). However, the genetic and phenotypic diversity within this species has not been systemically studied. Intra-species diversity can be characterized through various DNA-based methods. Many of these techniques that have been applied to biocontrol agents and *Cryptococcus* spp. do not involve sequencing, such as BOX-PCR, ERIC-PCR, amplified ribosomal DNA restriction analysis (McSpadden Gardener et al., 2000), terminal restriction fragment analysis (McSpadden Gardener, 2004) and amplified fragment length polymorphisms (Litvintseva, Thakur, Vilgalys, & Mitchell, 2006). Sequencing of conserved genetic loci, individually (Fritze, 2004; Samuels, 2006) or collectively as multilocus sequence typing (MLST) (Frapolli, Défago, & Moënne-Loccoz, 2007; Litvintseva et al., 2006) and multilocus microsatellite typing (Hanafy et al., 2008), was used when inference of phylogenetic relationships is desired. MLST was considered as a superior method for genotyping the human-pathogenic *Cryptococcus* spp., such as *C. neoformans* and *C. gattii*, owing to its high discriminating power and experimental reproducibility (Meyer et al., 2009). The comparison of multiple gene genealogies, as an outcome of MLST, is also the central theme of phylogenetic species recognition (Taylor et al., 2000). Phenotypic diversity among biocontrol agents has been investigated in a high-throughput fashion using Biolog assays (McSpadden Gardener et al., 2000).

The objectives of this study were (i) to genetically characterize 3C and 12 other *C. flavescens* strains through MLST, (ii) to determine their patterns of carbon source utilization and chemical sensitivity and (iii) to evaluate their biocontrol efficacy in regards to genotype. This research will provide insight into the relationships between genotype and phenotype of *C. flavescens* and guide marker development for distinguishing groups of *C. flavescens* with different genetic and phenotypic characteristics.

**Materials and Methods**
Strains

_Cryptococcus flavescens_ strain OH182.9_3C (3C) was obtained from the Northern Regional Research Laboratory (NRRL) under accession number NRRL Y-50378. Twelve other _C. flavescens_ strains were also obtained from NRRL under accession numbers Y-1401^T (= CBS 942^T), Y-7372, Y-7373, Y-7374, Y-7375, Y-7376, Y-7377, Y-7379, YB-328, YB-601, YB-602 and YB-744, respectively. The information on strain isolation was summarized in Table 9. All the strains were maintained as glycerol stocks at -80°C for long term storage.

Identification of MLST loci and choice of primers

The target genes of MLST were identified from 3C draft genome (GenBank accession numbers CAUG01000001 to CAUG01000712) through blastn and blastx search against NCBI Nucleotide collection (nr/nt) database. The boundaries of exons and introns were defined through Augustus prediction (Stanke et al., 2004) and alignment to _C. neoformans_ JEC21 (GenBank accession numbers AE107341 to AE107356). Primers used for MLST were adopted from previous studies (White, Bruns, Lee, & Taylor, 1990) or designed using software Primer 3 (Rozen & Skaletsky, 2000) (Table 10).

DNA isolation, PCR and sequencing

Genomic DNA of 3C and the 12 other _C. flavescens_ strains were extracted from 2-day old cultures grown on 1/5 Tryptic Soy Agar using the PowerSoil™ DNA Isolation Kit (MO BIO Laboratories, Inc., California, USA) following the manufacturer’s instruction. Each PCR reaction had a total volume of 25 µl, consisting of 2.5 µl genomic DNA of a _C. flavescens_ strain, 0.2 µl solutions of each primer (100 µM), 2.5 µl mixture of four deoxynucleoside triphosphates (2 mM each, mixed at 1:1:1:1 ratio), 1.8 µl 25 mM MgCl₂, 0.3 µl GoTaq® DNA polymerase (5 U/µl), 5 µl 5X Colorless GoTaq® Flex Buffer (Promega Corporation, Wisconsin, USA) and DNA-free PCR water as the rest of the volume. Thermal cycling program of 95°C for 7 min, then 33 cycles of 94°C for 1 min and 72°C for 2 min, followed by 72°C for 8 min was run on Peltier Thermal Cycler
PTC-200 (MJ Research, Inc., Massachusetts, USA). Both strands of amplification products were sequenced by the Molecular and Cellular Imaging Center at The Ohio State University (Wooster, Ohio, USA). Sequences were edited and concatenated using software Geneious 6.1.5 (created by Biomatters, available from http://www.geneious.com/). The majority of processed sequences were in high quality, where greater than 95% of the bases had Phred quality scores > 40. The few lower-quality bases were manually examined and base-called with enough confidence that they did not compromise the accuracy of phylogenetic analysis.

**Phylogenetic analysis**

Processed sequences were aligned and sequence identities were calculated using ClustalW2. Phylogenetic trees were generated using Maximum Likelihood method with General Time Reversible model in MEGA 5 (Tamura et al., 2011). Bootstrap test of 1,000 replicates were conducted to determine the reliability of consensus trees. To evaluate the effect of using different tree building methods, trees were also generated for all the relevant alignments using Neighbor Joining method in MEGA 5 and Bayesian inference in MrBayes v3.2.0 (Huelsenbeck & Ronquist, 2001), resulting in equivalent tree topologies to those by Maximum Likelihood method (data not shown).

**Biochemical assay**

Profiles of carbon source utilization and chemical sensitivity were generated using Biolog GEN III Microplates (Biolog, Inc., California, USA). Two-day-old *C. flavescens* colonies grown on 1/5 Tryptic Soy Agar were suspended in 120 µl of sterile double distilled water. Optical density of cell suspensions was measured at 595 nm using an ELx800 Universal Microplate Reader (BIO-TEK Instrument, Inc., Vermont, USA). 40 to 100 µl of cell suspension in water was added to 10 ml inoculation fluid IF-A (Biolog, Inc., California, USA) and mixed well by vortexing. 100 µl of cell suspension in IF-A (3 × 10⁵ to 4 × 10⁵ cells) was added to each well of a Biolog Gen III Microplate. Assay plates were incubated at room temperature (23 to 25°C) in the dark. Color change of assay wells, representing *C. flavescens* growth and cell respiration, was measured as
OD$_{595}$ by the ELx800 Universal Microplate Reader after 5 to 19 days of incubation. Two experiments were set up: 12 strains (3C, Y-7372, Y-7373, Y-7374, Y-7375, Y-7376, Y-7377, Y-7379, YB-328, YB-601, YB-602 and YB-744) were used in Experiment One; four strains (Y-1401T, 3C, Y-7372 and YB-601) were used in Experiment Two.

Statistical analyses were conducted on OD$_{595}$ data in Minitab 15 (Minitab Inc., Pennsylvania, USA). Cluster Variables analysis was performed on the data from all the assay wells as a whole using Correlation Coefficient Distance and a variety of Linkage methods including Average, Complete, McQuitty, Single and Ward. To avoid mistaking run-to-run variation for real phenotypic difference among strains, the minimum pair-wise similarity percentage between the two independent runs of the same strain was used as the lower similarity threshold among the strains forming one distinct group. Analysis of Variance (ANOVA) was performed on the data of individual assay wells to compare distinct groups using General Linear Model with an α value of 0.01, 0.03 or 0.1 for defining significant difference in $F$ test. In Experiment One, the data from 7 days of incubation in Run I and 10 days in Run II (referred as short incubation) were combined and those from 12 days in Run I and 15 days in Run II (referred as long incubation) combined for ANOVA. In Experiment Two, the data from the same days of incubation in both runs were combined for ANOVA.

**Biocontrol assays**

Macroconidia of *F. graminearum* isolate Z-3639 were produced on clarified V8 juice agar (CV8 agar) under a regime of 12 h/day fluorescent light for seven days at 25°C (Schisler et al., 2002). Colonized plates were flooded with weak PO$_4$ buffer to obtain conidial suspensions of the pathogen (Schisler et al., 2002). Biomass of all *C. flavescens* strains except Y-7374, Y-7375, Y-7376, Y-7377, Y-7379 was initiated on plates of one-fifth strength Tryptic soy broth agar (TSBA/5). After 24 h incubation at 28°C, cells were used to inoculate 50 ml of semi-defined complete liquid medium (SDCL; (Slininger et al., 2010)) in 250-ml Erlenmeyer flasks (OD= 0.10, A620). Flasks were incubated in a shaker incubator (250 rpm, 25°C, 24 h) to produce cells in log growth that were used to
inoculate cultures (50 ml SDCL in 250-ml flask, OD = 0.10, A620) that were used in experiments after incubation for 48 h.

Each experiment to determine strain efficacy was conducted in a climate-controlled greenhouse with temperatures that ranged from 17 to 20°C at night and 25 to 28°C during the day. Natural sunlight was supplemented with high-pressure sodium lights for 14 h/day. Each experimental unit consisted of two plants of hard red spring wheat cultivar Norm grown in a 19-cm-diameter plastic pot. Each pot contained air-steam pasteurized (60°C for 30 min) potting mix (Terra-lite Redi-earth mix, W.R. Grace, Cambridge, MA) and plants were grown in a growth chamber (25°C, 14-h photoperiod, 600 μmol/[m²/s]) for 7-8 weeks prior to transfer to greenhouse benches. Pots were fertilized after 1 week and weekly thereafter with 50 ml of a solution containing 1.25 g/liter Peters 20-20-20 (Grace-Sierra Horticultural Products, Milpitas, CA) and 0.079 g/liter iron chelate (Sprint 330, Becker Underwood, Inc., Ames, IA). Approximately one week after transferring plants to the greenhouse, wheat heads were inoculated at anthesis by spraying with 25% freshly harvested C. flavescens cells (~5 x 10⁷ CFU/ml) in PO₄ buffer and a final concentration of 0.036% Tween 80 (Sigma Chemical Co., St. Louis, MO). For each experiment, 50 mL of inoculum of each strain was used to inoculate six plants representing a total of 12 to 15 heads. Heads were then challenged immediately by spraying 12 mL of a conidial suspension of F. graminearum (3 x 10⁴ conidia/ml) in PO₄ buffer with 0.036% Tween 80. Wheat heads inoculated first with the PO₄ buffer/tween 80 solution and then only with F. graminearum conidia in PO₄ buffer/tween 80 served as a pathogen only control. Uninoculated pots of wheat were used to ensure pathogen inoculum did not spread between treatments. Inoculated plants were misted lightly with distilled water and incubated in a plastic humidity chamber at 17 to 20°C at night and 25 to 28°C during the day for 3 days before being transferred to greenhouse benches. Treatments were arranged in a completely randomized design. Fusarium head blight severity was visually estimated using a 0 to 100% scale (Stack & Mullen, 1995) at 10 and 16 days after inoculation. All greenhouse experiments were conducted five times. Non-parametric statistical test by the Kruskal-Wallis method was performed in Minitab 15.
Results

Evidence for two distinct genotypes of *C. flavescens*

Six loci – rDNA-ITS, the partial sequences of four protein-encoding genes and cs22 were sequenced and aligned to infer phylogenetic relationship among the *C. flavescens* strains (Table 10). These loci were probably unlinked since they were from six different scaffolds in the draft genome of 3C and their homologs were located on six different chromosomes of *C. neoformans* JEC21 (GenBank accession numbers AE107341 to AE107356), respectively. The rDNA-ITS sequences of all 13 strains were 100% identical. The other five loci revealed divergence within the species (thus termed variable loci), resulting in two clades (3C genotype and non-3C genotype) supported by a bootstrap value of 100 (Figure 8). For the four protein-encoding genes, percentage of identical sites in alignment ranged from 95.8% to 100% within each genotype and 89.5% to 95.1% between the two genotypes, where the Chitin Synthase 1 and β-tubulin loci showed the highest and lowest levels of divergence, respectively (Table 11). Locus cs22 is the 3’ end of a homolog to the terminal exon of a chitin synthase 5 gene in *C. neoformans* JEC21 (83 bp) plus its downstream anonymous region (724 bp) identified from the draft genome of 3C. The anonymous region is possibly intergenic because no reasonable coding sequences were predicted within it by Augustus and that neither discontiguous megablast nor blastx search against the NCBI Nucleotide Collection (nr/nt) returned any hit for it. The sequence difference between the two genotypes at cs22 was much lower than those for the four protein-encoding genes (Table 11), partially due to a 36-bp insertion (in 3C type) or deletion (in non-3C type) in the anonymous portion (data not shown). Interestingly, the sequence difference within each genotype at cs22 was medium to high instead of low as well compared to the four protein-encoding genes (Table 11).

The genotype split was consistent across all the five variable loci (Figure 9). There were two to six strains of each genotype showing 100% pairwise identity to one another at each of the five loci. In addition, all the corresponding sequences were 100% identical.
within each of the following groups: (i) 3C, Y-1401\(^T\), and Y-7373; (ii) YB-601 and YB-602; (iii) Y-7372, Y-7374, Y-7375 and (iv) YB-7379 and YB-744. The minimum pairwise identities for each locus were the same as or very similar to the percentage of identical sites in alignment. Y-7377 was divergent from the rest of the 3C-type strains by pairwise identities of 95.8%, 99.0% and 99.3% for three loci – chitin synthase 1, heat shock protein 70kDa and cs22 (Figure 9B, Figure 9D and Figure 9E; Table 11).

**Taxometric analysis on biochemical properties of C. flavescens strains shown in Biolog assay**

While sharing many similarities in response to the Biolog GEN III substrates, 12 C. flavescens strains (not including Y-1401\(^T\)) could be split into three measurably different groups, depending on incubation time. Common to all isolates was the good assimilation of 31 out of the 71 assayed carbonic compounds (Table 12, Row 1) and poor assimilation of 22 others (Table 12, Row 2). Additionally, all the 12 strains were resistant 18 out of 23 types of chemical stress (Table 12, Row 1) and sensitive to Tetrizolium Violet. All the 7 strains of non-3C genotype formed a distinct group from the 5 strains of 3C genotype after both short and long periods of incubation (A and B). At the end of the long incubation period, the basis of this grouping was found to come from difference in utilization of 11 carbon sources and sensitivity to 3 compounds (Table 12, Row 3). The two genotypes are thus considered two subspecies owing to the support from the phenotypic grouping by biochemical properties.

Interestingly, 3C-type strains showed different subgrouping patterns depending incubation time (A vs. B). YB-601 and YB-602 were quite distinct from all the other 10 strains after short incubation (A) while clustered together with the rest of 3C-type strains after long incubation (B). One of the contributors to this discrepancy is probably the unique utilization patterns of several carbon sources by YB-601 and YB-602, such as Tween 40, L-alanine, D-trehalose and L-malic acid. OD\(_{595}\) of these four assay wells were significantly lower for YB-601 and YB-602 than for the rest of 3C-type strains (\(P < 0.01\) for F test) after short incubation while not significantly different among the two subgroups of 3C type (\(P > 0.1\) for F test) after long incubation in Experiment One. This
indicated that YB-601 and YB-602 assimilated these carbon sources more slowly, resulting in a lagged growth compared to the other 3C-type strains. This phenotypic difference is consistent with the divergence of YB-601 and YB-602 from the three 3C-type strains in the phylogeny of elongation factor 1 (Figure 9C). After long incubation, Y-7373 and Y-7377 formed a phenotypically distinct group from 3C, YB-601 and YB-602 by a similarity slightly lower than the threshold (B). This is partially consistent with the divergence of Y-7377 from the other three strains at the loci of chitin synthase 1, heat shock protein 70kDa and cs22 (Figure 9B, Figure 9D and Figure 9E). The use of other Linkage methods (Average, McQuitty, Single and Complete) than Ward in Cluster analysis resulted in different topologies and similarity magnitudes (data not shown). However, the aforementioned clustering patterns were always demonstrated regardless of Linkage method.

Type strain Y-1401^T showed distinct biochemical properties from 3C, YB-601 (both 3C genotype) and Y-7372 (non-3C genotype), regardless of incubation time (Figure 10). This distinction is probably due to poorer growth of Y-1401^T than the other three strains, with statistical significance in 27 out of 71 carbon source utilization assays and 16 out of 23 chemical sensitivity assays (Table 12, Row 4). The only exception was that Y-1401^T utilized citric acid slightly better than the other strains by the end of 19 days’ incubation (P = 0.086 for F test).

Biocontrol efficacy of the two subspecies of C. flavescens suppressing Fusarium Head Blight (FHB)

Representative strains from the two genotypes (3C type and non-3C type) were assessed for FHB suppression on wheat in five separate greenhouse bioassays. Reduction in disease severity caused by pre-inoculation with C. flavescens strains were observed in four of five independent trials. In two of those trials, the observed pattern was statistically significant (P < 0.10, Kruskal-Wallis test) after 16 days of incubation (data not shown). In total, the 3C-type strains displayed greater biocontrol activity than the non-3C-type strains in three of five independent trials, but given the high degree of variation of disease scores that are typical of such trials, the individual run results were not statistically
significant (data not shown). However, overall, significant reductions in disease severity due to inoculation by *C. flavescens* were observed at both 10 and 16 days post-inoculation \((P = 0.11 \text{ and } 0.03, \text{ respectively, Kruskal-Wallis test; Figure 12})\). Most significantly, 3C type reduced disease severity relative to the negative control after 16 days of incubation \((P < 0.10 \text{ Kruskal-Wallis test})\). Non-3C type also appeared to reduce disease severity, but the pattern was not statistically significant. As a group, 3C type reduced average disease severity at 10 days (16% versus 23% for the negative control) and 16 days (46% versus 64% for the negative control) post inoculation. Non-3C type reduced average disease severity (to 17% and 51% at 10 and 16 days, respectively), but not as dramatically as 3C type. These data demonstrate that both genotypes of *C. flavescens* have the capacity to control Fusarium head blight, and of the two, 3C type is more effective at controlling FHB than non-3C type.

**Discussion**

Here we present the first detailed genotypic and phenotypic characterization of *C. flavescens* species and note the occurrence of two distinct subspecies. The only previously available sequences of *C. flavescens* were the ribosomal DNAs and internal transcribed spacers of type strain Y-1401\(^T\) (= CBS 942\(^T\)) and several other strains (Takashima et al., 2003). All rDNA-ITS sequences obtained in this study were 100% identical to the published sequence of Y-1401\(^T\) (GenBank accession number AB035046). Since a strain with 99.0% identity to Y-1401\(^T\) in ITS region has been identified as *C. flavescens* (Sugita et al., 2000; Takashima et al., 2003), all the other 12 strains in this study would typically be classified as *C. flavescens* as well based on ITS genotype alone. The genotype split shown by the five variable MLST loci was consistent with the results of a previous quantitative PCR (qPCR) assay amplifying a 150-bp region within the same heat shock protein 70kDa locus used in MLST of this study (X. Rong, P. A. Paul, D. A. Schisler, B. B. McSpadden Gardener, submitted for publication). The qPCR assay differentiated the tested *C. flavescens* strains into two groups, 3C like and non-3C like,
based on difference in threshold cycles that resulted from sequence difference in primer binding regions. The 3C like strains actually belong to 3C genotype and non-3C like to non-3C genotype. Within 3C genotype, the 100% identity of Y-1401\(^T\) to 3C and Y-7373 at the five variable MLST loci indicates that Y-1401\(^T\), though isolated in Japan many decades earlier, is closely related to the two strains isolated more recently in the USA (Khan et al., 2001; Kurtzman, 1973; Saito, 1922). Thus, the species appears to be genetically stable on a decadal time scale.

In Biolog assays, the fact that all the strains of non-3C genotype formed a distinct cluster than 3C-type strains shows that genotype is indicative of biochemical properties for \textit{C. flavescens}. The utilization results of carbon sources were consistent most of the time between previous studies and this study (Table 12, Row 5). The only minor inconsistency was glycerol. The results were previously negative for Y-1401\(^T\) (CBS-KNAW Fungal Biodiversity Centre [http://www.straininfo.net/strains/103522/browser], as of May 27, 2013) or variable (Fonseca et al., 2011; Takashima et al., 2003). However, it was positive or near positive for utilization by all the 13 strains in our study (OD\(_{595}\) = 0.589 for Y-1401\(^T\) and OD\(_{595}\) > 0.7 for the other 12 strains after longest incubation). Though Y-1401\(^T\) showed minimal divergence from most of the 3C-type strains, its Biolog results were highly distinct due to generally poorer growth. It is unknown whether this pattern resulted from geographic/habitat difference or sub-culturing/storage of strains over the years. Carbon source utilization has been previously related to recovery of biocontrol agents against Fusarium head blight (FHB), where utilization of tartaric acid by microorganisms was positively biased towards the ability to reduce FHB damages (Khan et al., 2001). Tartaric acid can be utilized by OH182.9 (progenitor of 3C) (Khan et al., 2001) but not included in the Biolog assays used in this study. Whether the Biolog assays have any other indications of \textit{C. flavescens} biocontrol efficacy needs further investigation.

Both genotypes of \textit{C. flavescens} showed biocontrol efficacy against FHB, which indicates the species as a whole has biocontrol potential. 3C genotype performed slightly better than non-3C type when combining data from five independent runs. However, the results were variable when splitting data by individual strain and/or run. This level of
variation is common in biocontrol efficacy assays against FHB in general (D. A. Schisler, personal communication). In conclusion, genotype is weakly predictive of biocontrol efficacy under the assay conditions in this study.

Several strains were previously designated as two mating types based on a mating test where mating responses included formation of conjugation tubes and hypha growth from fused tubes in mixed culture of strain pairs (Kurtzman, 1973). The strains of the opposite mating types are able to mate (in the lab) while those of the same mating type are not. In that study, Y-7372, Y-7374, Y-7375, Y-7376 and Y-7379 were designated as mating type α; while Y-7373 and Y-7377 as mating type a. Interestingly, all the five α-strains were shown as non-3C type while both a-strains as 3C type by MLST in our study. The genetic divergence between the two mating types indicated minimal sexual recombination among the studied strains in the environment. The fact that the strains of the same MLST genotype belonged to the same mating type in this study is consistent with a previous MLST study on *C. neoformans* var. *grubii*, where more than 40 MLST types and only two mating types were identified (Litvintseva et al., 2006). Thus, it is possible that within a *Cryptococcus* species, the strains of different mating types tend to be of different genotypes at other conserved loci.

Genetic markers have been used to assist discovery of novel biocontrol isolates and to study the geographic distribution of biocontrol agents (Benitez & McSpadden Gardener, 2009; McSpadden Gardener, Gutierrez, Joshi, Edema, & Lutton, 2005). The MLST markers developed in this study can be used to identify biocontrol-relevant substrates positive for *C. flavescens* presence so as to minimize the efforts in isolating novel *C. flavescens* biocontrol agents. MLST analysis can also be performed on more *C. flavescens* strains recovered from wheat fields at various geographic locations to comprehensively characterize the natural occurrence of this species in biocontrol-relevant habitats. Heterosis, where hybrid progeny shows superior biological quality over parents, has been utilized in plant breeding to produce hybrid cultivars with improved performance such as higher yield and disease resistance (Birchler, Yao, Chudalayandi, Vaiman, & Veitia, 2010). Crossing *C. flavescens* isolates of compatible mating types can be conducted to investigate whether hybrid progeny shows increased biocontrol efficacy
against FHB. The carbon source utilization and chemical resistance patterns shown in Biolog may also be used to develop *C. flavescens*-specific media aiding the selective recovery of *C. flavescens* as potential biocontrol agents.
<table>
<thead>
<tr>
<th>NRRL Accession No.</th>
<th>Isolation substrate</th>
<th>Isolation location</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-1401 (Type strain)</td>
<td>Atmosphere</td>
<td>Japan</td>
<td>(Saito, 1922)</td>
</tr>
<tr>
<td>Y-50378 (OH182.9_3C)</td>
<td>Wheat anther</td>
<td>Ohio or Illinois, USA</td>
<td>(Khan et al., 2001)</td>
</tr>
<tr>
<td>Y-7372</td>
<td>Surface of dent corn kernels of moisture 15%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>Y-7373</td>
<td>Surface of dent corn kernels of moisture 15%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>Y-7374</td>
<td>Surface of dent corn kernels of moisture 30%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>Y-7375</td>
<td>Surface of dent corn kernels of moisture 12%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>Y-7376</td>
<td>Surface of dent corn kernels of moisture 17%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>Y-7377</td>
<td>Surface of dent corn kernels of moisture 20%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>Y-7379</td>
<td>Surface of dent corn kernels of moisture 25%</td>
<td>Central Illinois, USA</td>
<td>(Kurtzman, 1973)</td>
</tr>
<tr>
<td>YB-328</td>
<td>Soil</td>
<td>Hatcha NRRL catalog</td>
<td></td>
</tr>
<tr>
<td>YB-601</td>
<td>Enzyme solution contaminant</td>
<td>R. Dimler, NRRL³ NRRL catalog</td>
<td></td>
</tr>
<tr>
<td>YB-602</td>
<td>Enzyme solution contaminant</td>
<td>R. Dimler, NRRL³ NRRL catalog</td>
<td></td>
</tr>
<tr>
<td>YB-744</td>
<td>Unknown</td>
<td>Nelson, Iowa State University, Ames, Iowa³ NRRL catalog</td>
<td></td>
</tr>
</tbody>
</table>

*This information is from the “Source” field of NRRL catalog and may not reflect the actual geographic location of isolation, which could not be identified from available records.

Table 9. *C. flavescens* strains used in this study
<table>
<thead>
<tr>
<th>Primer name</th>
<th>Sequence (5’ – 3’)</th>
<th>Target locus</th>
<th>Scaffold accession number</th>
<th>Number of bp(^a)</th>
<th>Exon/Intron(^b)</th>
<th>Whole gene</th>
<th>Amplified</th>
<th>Used in MLST</th>
<th>Number</th>
<th>Total bp(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITS4</td>
<td>TCCTCCGCTTATTGATATGC</td>
<td>18S(partial)-ITS1-5.8S-ITS2-25S(partial)</td>
<td>CAUG01000137</td>
<td>Longer than the scaffold</td>
<td>553</td>
<td>531</td>
<td>1/2</td>
<td>247/284</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ITS5</td>
<td>GGAAGTAAAAGTCGTAACAAGG</td>
<td>β-tubulin</td>
<td>CAUG01000343</td>
<td>1641</td>
<td>585</td>
<td>534</td>
<td>4/3</td>
<td>379/155</td>
<td></td>
<td></td>
</tr>
<tr>
<td>btb.m1-2F</td>
<td>GGGCGGTTGGAGGGAGAAGATCACG</td>
<td>Chitin synthase 1</td>
<td>CAUG01000683</td>
<td>3840</td>
<td>356</td>
<td>332</td>
<td>2/1</td>
<td>229/103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS24.m2R</td>
<td>CGAACGAGGTGAATACGGTGC</td>
<td>Elongation factor 1</td>
<td>CAUG01000040</td>
<td>2012</td>
<td>658</td>
<td>628</td>
<td>3/2</td>
<td>450/178</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS24.m2F</td>
<td>ACAGGCTTGAGGGAGAAGATCACG</td>
<td>Heat shock protein 70kDa</td>
<td>CAUG01000237</td>
<td>2261</td>
<td>719</td>
<td>698</td>
<td>5/4</td>
<td>508/190</td>
<td></td>
<td></td>
</tr>
<tr>
<td>h31.m1F</td>
<td>ATGTACGCCGCTACCCGATCTCTTC</td>
<td>cs22 (chitin synthase 5 plus downstream anonymous region)</td>
<td>CAUG01000018</td>
<td>6412(^c)</td>
<td>91, 771(^d), 83, 724(^d)</td>
<td>1/0(^e), 83, 0(^f)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h31.m1R</td>
<td>CGACAAAGGCAAATTTGGGATGG</td>
<td>Chitin synthase 7</td>
<td>CAUG01000482</td>
<td>3183</td>
<td>420</td>
<td>399</td>
<td>2/1</td>
<td>353/46</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS22.m2F</td>
<td>AAGAAGGGAGTAGGAGACAGCTCG</td>
<td>Heat shock protein 70kDa</td>
<td>CAUG01000166</td>
<td>2261</td>
<td>831</td>
<td>800</td>
<td>4/3</td>
<td>653/147</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS22.m2R</td>
<td>ACTTGGTTCGCTTGGTAGATGG</td>
<td>Heat shock protein 70kDa</td>
<td>CAUG01000224</td>
<td>2080</td>
<td>834</td>
<td>772</td>
<td>5/4</td>
<td>546/226</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) The numbers were precise for 3C but approximate for other strains since the some strains have insertion or deletion (less 1% of the total numbers).

\(^b\) Information represents the portions used in MLST. The rDNA (18S, 5.8S, 25S) and ITS regions were counted here as exon and intron, respectively.

\(^c\) Information represents chitin synthase 5 only.

\(^d\) Information represents chitin synthase 5 and its downstream anonymous region, respectively.

Table 10. Primers and target loci for the MLST analysis on *C. flavescens* strains
<table>
<thead>
<tr>
<th>Target locus</th>
<th>% identical sites$^a$ among …</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3C type$^b$</td>
</tr>
<tr>
<td>18S-ITS1-5.8S-ITS2-25S</td>
<td>100</td>
</tr>
<tr>
<td>β-tubulin</td>
<td>100</td>
</tr>
<tr>
<td>Chitin synthase 1</td>
<td>95.8</td>
</tr>
<tr>
<td>Elongation factor 1</td>
<td>99.7</td>
</tr>
<tr>
<td>Heat shock protein 70kDa</td>
<td>99.0</td>
</tr>
<tr>
<td>cs22 (chitin synthase 5 plus downstream anonymous region)</td>
<td>99.3</td>
</tr>
<tr>
<td>Concatenation of the 5 regions above (excluding rDNA-ITS)</td>
<td>99.0</td>
</tr>
<tr>
<td>Chitin synthase 7</td>
<td>99.7</td>
</tr>
<tr>
<td>Heat shock protein 70kDa (h22)</td>
<td>100</td>
</tr>
<tr>
<td>Heat shock protein 70kDa (h30)</td>
<td>98.7</td>
</tr>
</tbody>
</table>

$^a$ The percentage of columns in the alignment for which all the sequences were identical

$^b$ The strains classified as 3C type are 3C, Y-1401$^T$, Y-7373, Y-7377, YB-601 and YB-602; those classified as non-3C type are Y-7372, Y-7374, Y-7375, Y-7376, Y-7379, YB-328 and YB-744.

Table 11. DNA sequence identities among *C. flavescens* strains
Table 12. Biochemical characteristics of C. flavescens strains shown by OD$_{595}$ of Biolog assay (continued)

$^a$ Information in these rows is based on the Experiment One data combined from 12 days of incubation in Run I and 15 days of incubation in Run II. Y-1401$^T$ was not included.

$^b$ Information in this row is based on the Experiment Two data combined from 19 days of incubation in both Run I and Run II.

$^c$ The results from previous studies (Fonseca et al., 2011; Takashima et al., 2003) were indicated as “+” for positive growth, “−” for negative and “v” for variable in the parentheses following each chemical. In our study, OD$_{595}$ was $> 0.6$ for “+”, $< 0.4$ for “−” and a mixture of $> 0.6$ and $< 0.4$ for “v” by 19 days of incubation in Experiment Two Run II for Y-1401$^T$ and by 15 days of incubation in Experiment One Run II for the other 12 strains. OD$_{595}$ of Y-1401$^T$ for D-galactose utilization assay was 0.472.
### Table 12. Continued

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Carbon source</th>
<th>Chemical stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good growth of all 12 strains (OD&lt;sub&gt;595&lt;/sub&gt; &gt; 0.6), sorted by mean OD&lt;sub&gt;595&lt;/sub&gt; among strains in descending order&lt;sup&gt;a&lt;/sup&gt;</td>
<td>D-cellobiose, stachyose, D-trehalose, D-turanose, α-D-lactose, β-methyl-D-glucoside, γ-amino-butyric acid, D-mannitol, D-melibiose, D-raffinose, D-maltose, D-galactose, D-fructose, L-pyroglutamic acid, sucrose, L-glutamic acid, gentiobiose, L-rhamnose, L-malic acid, D-arabinol, L-arginine, D-gluconic acid, L-aspartic acid, L-galactonic acid lactone, L-alanine, D-mannose, acetic acid, D-galacturonic acid, α-D-glucose, glycy1-L-proline, L-serine</td>
<td>1% sodium lactate, 1% NaCl, potassium tellurite, pH 5, pH 6, aztreonam, nalidixic acid, lincomycin, rifamycin SV, vancomycin, minocycline, sodium butyrate, fusidic acid, D-serine, 4% NaCl, sodium bromate, troleandomycin</td>
</tr>
<tr>
<td>Poor to no growth of all 12 strains (OD&lt;sub&gt;595&lt;/sub&gt; &lt; 0.6), sorted by mean OD&lt;sub&gt;595&lt;/sub&gt; among strains in ascending order&lt;sup&gt;a&lt;/sup&gt;</td>
<td>α-Hydroxy-butyric acid, p-hydroxy-phenylacetic acid, quinic acid, propionic acid, β-hydroxy-D,L-butyric acid, formic acid, inosine, N-acetyl neuraminic acid, gelatin, D-serine, α-keto-butyric acid, D-glucose-6-PO&lt;sub&gt;4&lt;/sub&gt;, L-histidine, N-acetyl-D-galactosamine, D-lactic acid methyl ester, N-acetyl-β-D-mannosamine, pectin, N-acetyl-D-glucosamine, D-fructose-6-PO&lt;sub&gt;4&lt;/sub&gt;, citric acid, acetoacetic acid</td>
<td>Tetrazolium violet</td>
</tr>
<tr>
<td>Significant difference between the 3C and non-3C genotypes (P &lt; 0.01 for F test), sorted by P-value in ascending order&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Propionic acid, L-histidine, acetoacetic acid, glucuronamide, D-aspartic acid, D-malic acid, L-lactic acid, 3-methyl glucose, p-hydroxy-phenylacetic acid, pectin, glycy1-L-proline</td>
<td>Lithium chloride, Niaproof 4, tetrazolium blue</td>
</tr>
<tr>
<td>Significantly poorer growth of Y-1401&lt;sup&gt;T&lt;/sup&gt; than three other strains (P &lt; 0.03 for F test), sorted by P-value in ascending order&lt;sup&gt;b&lt;/sup&gt;</td>
<td>D-raffinose, α-D-glucose, D-melibiose, D-fructose, D-galactose, α-D-lactose, D-cellobiose, β-methyl-D-glucoside, dextrin, D-mannose, D-galacturonic acid, bromo-succinic Acid, L-aspartic acid, D-glucuronic acid, sucrose, mucic acid, L-alanine, Tween 40, L-rhamnose, glycerol, D-maltose, L-serine, D-fucose, D-salicin, L-arginine</td>
<td>D-serine, 1% NaCl, pH 5, sodium bromate, potassium tellurite, 1% sodium lactate, minocycline, fusidic acid, aztreonam, guanidine HCl, pH 6, nalidixic acid, sodium butyrate, troleandomycin, vancomycin</td>
</tr>
<tr>
<td>Consistent with previous studies&lt;sup&gt;c&lt;/sup&gt;</td>
<td>N-acetyl-D-glucosamine (+), D-cellobiose (+), D-galactose (+) (expect Y-1401&lt;sup&gt;T&lt;/sup&gt;), D-galacturonic acid (+), D-glucuronic acid (+), α-D-glucose (+), D-glucuronic acid (+), sucrose (+), D-raffinose (+), D-melibiose (+), L-lactic acid (v), α-D-lactose (+), D-maltose (+), β-methyl-D-glucoside (+), myo-inositol (+), D-salicin (+)</td>
<td>Not applicable</td>
</tr>
</tbody>
</table>

---

<sup>a</sup>Sorted by mean OD<sub>595</sub> among strains in ascending or descending order.

<sup>b</sup>Sorted by P-value in ascending order.

<sup>c</sup>Consistent with previous studies.
Figure 8. Genetic similarity among C. flavescens strains shown by the concatenated DNA sequences of β-tubulin, Chitin Synthase 1, Elongation Factor 1 and Heat Shock Protein 70 kDa.

The phylogenetic relationships among taxa were inferred by the Maximum Likelihood method with General Time Reversible model. Bootstrap values of > 60% (1,000 replicates) are shown next to the branches. Scale bar shows number of base substitutions per site. The two groups of C. flavescens strains split into two clades in were defined as 3C type and non-3C type, respectively.
Figure 9. *C. flavescens* strains consistently split into two groups shown in individual DNA trees (continued).

The trees were constructed based on the DNA sequences of β-tubulin (A), Chitin Synthase 1 (B), Elongation Factor 1 (C), Heat Shock Protein 70 kDa (D) and chitin synthase 5 plus downstream anonymous region (E). All the 3C-type strains were in the same clade (lower) while the strains of non-3C genotype in the other clade (upper). The phylogenetic relationships among taxa were inferred by the Maximum Likelihood method with General Time Reversible model. Bootstrap values of > 60% (1,000 replicates) are shown next to the branches. Scale bar shows number of base substitutions per site.
Figure 9. Continued

A (β-tubulin)  
Y-7375  
Y-7374  
Y-7372  
YB-328  
Y-7379  
Y-7376  
YB-744  
YB-601  
YB-602  
Y-7373  
Y-1401\textsuperscript{T}  
3C  
Y-7377

B (chitin synthase 1)  
Y-7376  
Y-7374  
Y-7372  
YB-328  
Y-7379  
YB-744  
Y-7377  
Y-1401\textsuperscript{T}  
3C  
YB-602  
Y-7373  
YB-601

C (elongation factor 1)  
YB-328  
Y-7376  
YB-744  
Y-7379  
YB-602  
Y-7377  
YB-601  
Y-1401\textsuperscript{T}  
3C  
Y-7373
Figure 9. Continued

D (heat shock protein 70 kDa)   E (chitin synthase 5 plus downstream anonymous region)
Figure 10. Cluster analysis on the carbon source utilization and chemical sensitivity patterns of C. flavescens strains at two time points (A and B) in Experiment One. Type strain Y-1401T was not included. (Continued)

Two independent runs of Biolog assays were performed for each strain. Due to the variation in assay conditions between the two runs, the data from 7 and 12 days of incubation in Run I were approximately equivalent to those from 10 and 15 days of incubation in Run II, respectively. Cluster Variables analysis using Correlation Coefficient Distance and Ward Linkage was conducted in Minitab 15 on the OD$_{595}$ measurements of assay plates. Each distinct group was defined by a similarity percentage greater than the minimum between two replicates of the same strain (*). Different bracket line patterns were used to differentiate 3C-type (homogenous dash) and non-3C type strains (alternative dash).
Figure 10. Continued

A

Continued
Figure 10. Continued

B

Strain, Run, Day

Similarity %
Figure 11. Cluster analysis on the carbon source utilization and chemical sensitivity patterns of type strain Y-1401^T and other selected *C. flavescens* strains at two time points (left and right) in Experiment Two.

Two independent runs of Biolog assays were performed for each strain. Cluster Variables analysis using Correlation Coefficient Distance and Ward Linkage was conducted in Minitab 15 on the OD_{595} measurements of assay plates. Each distinct group was defined by a similarity percentage greater than the minimum between two replicates of the same strain (*). All the strains were 3C genotype except Y-7372 which was non-3C type.
The assayed strains were 3C, Y-7373, YB-601 and YB-602, which were 3C type, as well as non-3C-type strains Y-7372, YB-328, and YB-744. Data collected at two time points post inoculation on average % FHB disease severity (percentage of infected area per head) of greenhouse-grown wheat inoculated with *Fusarium graminearum* were combined across five independent experiments. A total of 227, 158, and 69 heads were assayed for 3C type, non-3C type, and the negative control (NC), respectively. Error bars represent standard error of the means.
Chapter 5: Future research

The genome of 3C was sequenced and molecular markers were developed (Chapter 2) and used (Chapter 3 and 4) to quantify the abundance and to characterize the diversity of C. flavescens strains from the environment. The initial application of a qPCR assay based on one of the markers showed stochastic dispersal pattern of 3C-like C. flavescens in a wheat field throughout the growing season. Overwintering persistence and die-off on threshed grains during storage were also demonstrated by the qPCR. A multilocus sequence typing (MLST) scheme was developed as a combination of multiple markers. MLST analysis showed two genotypes congruent to mating types in the species of C. flavescens, which were supported by biochemical and biocontrol data. These set the stage for the future research directions proposed below.

Genome-wide discovery of genes involved in biocontrol of 3C against FHB

I hypothesize that (i) multiple genes in the 3C genome are involved in biocontrol against FHB, (ii) many of those genes will be differentially expressed depending on whether biocontrol is occurring and (iii) deletion of such genes may affect biocontrol efficacy. Protein encoding genes in the genome assembly of 3C were predicted \textit{ab initio} through Augustus using C. neoformans JEC21 as a model. Draft functional annotation of the genome was obtained through JGI automated annotation and blastx search against the NCBI database non-redundant protein sequences (nr). An initial attempt was made to identify 3C genes related to biocontrol. Whether those genes are truly functional genes
instead of pseudogenes, whether they indeed affect biocontrol and whether there are other genes involved in biocontrol needs further investigation. Therefore, the following research is proposed.

**Transcriptomic analysis of interactions among 3C, *Fusarium graminearum* (Fg) and wheat**

A growth chamber experiment can be performed with the following four inoculation treatments applied at flowering of wheat: 3C alone, Fg alone, 3C + Fg, and buffer. In addition, 3C cells from the inoculum can be sampled as a control. To simplify the conditions, a susceptible wheat cultivar and a highly virulent Fg isolate can be used in the initial experiments, but multiple wheat cultivars and Fg isolates with differential resistance and virulence, respectively, can be used in follow-up experiments. To better define the dynamics of the ecological interactions, wheat heads can be harvested at three time points between inoculation and harvest. There can be four biological replicates per treatment per time point. The technology of strand-specific RNA-Seq (Perkins et al., 2009) can be used to sequence the whole transcriptomes of participating organisms. The cDNA reads obtained from RNA-Seq can be mapped to the genomes of 3C, wheat and Fg, the latter two of which are publicly available at www.wheatgenome.org/ and www.broadinstitute.org/annotation/genome/fusarium_group/MultiHome.html, respectively. The differential expression of genes can be analyzed based on the read alignments and reconstructed transcriptomes using algorithms such as DESeq (Levin et al., 2010).

These analyses can help answer the following questions related to biocontrol mechanisms of 3C. 1) How do 3C transcriptomes differ *in vitro* versus *in planta*? 2) How will 3C transcriptome change when it is acting against FHB? 3) How is wheat transcriptome altered by 3C colonization, especially is host resistance to FHB induced? And 4) how can 3C alter Fg transcriptome with regards to pathogenesis?

I hypothesize that important biocontrol genes of 3C will show significant differential expression during colonization of wheat and/or co-existing with Fg. Attempts can be made to classify such genes into functional categories related to biocontrol, i.e. antibiosis,
competition, and induction of host resistance (Kim et al., 2011). Sequence homology-based annotation can provide some clues as to the functions of these genes, help with the further selection of 3C genes for gene knockout analyses and suggest approaches to manipulating the production and/or application of 3C-based biopesticide to improve its efficacy. In addition, some wheat and Fg genes are also expected to be differentially transcribed depending on the presence of 3C. Specifically, it is expected that 3C alters the physiology of wheat heads and Fg, the pathogenesis-related interactions between the host and the pathogen and possibly the expression of their pathways related to DON accumulation. For example, we can look out for wheat genes involved in induced systemic resistance against diseases, which is usually triggered by non-pathogenic plant-associated microorganisms, as well as Fg genes related to infection and disease development on wheat.

**Gene knockout analyses on the 3C genes potentially involved in biocontrol**

Gene knockout analyses can be performed on the genes that showed significant differential expression depending on the treatments. It is possible that some biocontrol-related genes are expressed constitutively and not greatly affected by those proposed treatments. Therefore, knockout analyses can also be performed on the genes identified in genomic analysis that were highly homologous to the known biocontrol-related genes, especially those that did not always show differential expression in biocontrol-relevant scenarios.

Gene knockout analyses have been conducted on *Cryptococcus neoformans*, which belongs to the same genus as 3C but is a different species. Specific exogenous DNA can be delivered into *C. neoformans* (Davidson et al., 2000), and RNA interference has been used in *C. neoformans* to silence the genes of interest (Bose & Doering, 2011; Liu, Cottrell, Pierini, Goldman, & Doering, 2002). Preliminary experiments can be performed to test whether these methods can successfully knock out 3C genes of interest. Then the research can proceed with the approach that provides a better balance of productivity versus cost for the 3C system.
To examine the effects of gene knockout on the biocontrol activities of 3C, a series of bioassays can be conducted. To test for changes in efficacy against FHB, a growth chamber experiment can be performed where 3C wild type (WT) and knockout transformants are applied separately to wheat inoculated with Fg. Data can be collected on disease incidence, disease severity and 100 kernel weight of wheat (Khan et al., 2004; Schisler et al., 2002). Since biocontrol of one organism typically involves the expression of multiple mechanisms (Kim et al., 2011), loss of one mechanism does not always result in detectable changes in the overall biocontrol efficacy. Therefore, other bioassays can be performed to test the loss of specific functions which may not be manifested as changes in overall efficacy. For example, to assess knockouts in 3C genes potentially involved in direct antibiosis, in vitro antibiotic assay can be conducted by co-culturing 3C WT and transformants with Fg. To assess 3C genes involved in inducing host defense pathways, changes in wheat defense gene expression can be assessed using qRT-PCR following 3C treatment. And, to assess the importance of catabolism genes involved in competition, comparison can be made between the abilities of 3C WT and transformants to grow in the presence and absence of Fg on specific compounds produced by wheat that may be inductive to Fg pathogenesis.

In addition, the quantitative PCR assay can be used to evaluate association of the population levels with the biocontrol efficacy of 3C WT versus mutants with impaired biocontrol. Quantification of Fg population during bioassay can also be performed to investigate its correlation with 3C WT vs. mutant populations. This type of quantification work can help understand how the gene of interest is related to biocontrol.

Reduction in 3C transformants’ efficacy against FHB and/or its biocontrol-related metabolisms can demonstrate the involvement of corresponding genes in the biocontrol functions of 3C and provide guidance to improving efficacy by manipulating formulation and application. Specifically, if those genes appear to be involved in the secretion of antibiotic metabolites against Fg, 3C-based biopesticides should be formulated and applied in the ways that maximize the production of those antibiotics and create the micro-environments where the antibiotics can exert their maximum effects against the pathogens. Alternatively, if those genes appear to be involved in competition of nutrients
with Fg, then the formulation and application should minimize the availability of the nutrients for which 3C will compete with Fg and maximize 3C’s competitive power in itself. Or, if the genes act primarily through induction of host resistance pathways against Fg, then the formulation and application should condition both 3C and wheat into the states where host resistance can be induced to the maximal extent upon Fg infection. These results can help us understand the biocontrol mechanisms of flower-colonizing yeasts and suggest approaches to improve the efficacy of biopesticides based on these organisms.

**Characterizing the diversity and abundance of natural *C. flavescens* populations in wheat and barley**

**Validity of using developed qPCR and MLST primers for future studies**

Previous research showed that 3C was isolated from wheat (Khan et al., 2001) and that the distribution of biocontrol agents depends on geographic locations (McSpadden Gardener et al., 2005). Therefore I hypothesize that (i) natural *C. flavescens* populations are detectable in wheat and possibly barley and (ii) the diversity and abundance of such populations vary across United States. As described in Chapter 3, we developed a real-time quantitative PCR assay to monitor the dispersal and survival of 3C in the field and on threshed grains during storage. During Multilocus Sequence Typing (MLST), we identified 8 conserved loci in *C. flavescens* genome that are polymorphic between 3C and non-3C genotypes, which we can use to genetically classify *C. flavescens* isolates.

Even though the qPCR primer pair h31.2 was not able to differentiate 3C from some other 3C-type strains, doubt about the value of this qPCR assay due to primer specificity should be minimal. The DNA sequences in those 8 characterized loci were 100% identical between 3C and Y-7373 and more than 99.9% identical between 3C and other 3C-type strains except Y-7377. This indicates it is not very possible to discover other loci that are both polymorphic enough to distinguish 3C from all the other 3C-type strains and
conserved enough to be consistently PCR-amplified for detection. Therefore, the current qPCR primers are satisfactory in terms of specificity and will be proper to use for quantifying the abundance of 3C population across the production systems of the host crops it can protect.

However, the existence of natural 3C-type *C. flavescens* populations will confound the interpretation about whether the qPCR-detected populations result from the reproduction of artificially applied 3C inoculum. Therefore, a quantification of natural *C. flavescens* population is necessary to understand to what extent the natural populations will interfere with quantification of 3C population arising from applied inoculum. We can use the developed qPCR assay for this quantification.

The MLST primers can also be used to determine whether natural *C. flavescens* populations of non-3C genotype are present in crop production systems. To address the concern about the specificities of the MLST primers to *C. flavescens* species, blastn search was performed using the MLST primers as query against the genome of *C. neoformans* JEC21 (accession numbers AE017341 to AE017356). Seven out of the 8 primer pairs showed higher sequence difference than 9% of the total nucleotides in the forward plus reverse primers compared to the hits in JEC21. The only exception was h30.m1-2F plus h30.m1R, which showed 8% difference from JEC21 hits. This means that none of the 7 MLST primer pairs is able to amplify *C. neoformans* DNA more efficiently than the qPCR primers amplify non-3C genotype, since each of the qPCR primers has 9% sequence difference from its binding site in non-3C genotype. It indicates that all of the 7 MLST primer pairs are able to differentiate *C. neoformans* from *C. flavescens*. No genome sequences are available for the *Cryptococcus* species that are more closely related to *C. flavescens* than *C. neoformans* is.

With the qPCR and MLST tools available, a study on characterizing the diversity and abundance of natural *C. flavescens* populations in wheat production system is proposed. Since barley is also a host of Fusarium head blight and 3C may have biocontrol efficacy on barley, the characterization and quantification will be conducted on barley as well.
Characterizing the diversity and abundance of natural *C. flavescens* populations using the developed qPCR and MLST primers

To represent the natural *C. flavescens* populations from various geographic regions, wheat and barley samples can be collected from multiple states across United States, where epidemics of Fusarium head blight have been reported in recent years. Multiple fields of different cultivars of the two crops in different regions within each state can be sampled, and samples can be collected from different parts of each field. Heads and the stem-leaf portion of the plants can be collected as two types of samples. Data on FHB disease damages can also be collected if interesting patterns are observed. Sampling can be conducted at two time points – during anthesis and shortly before harvest. Threshed grains and ground flours can also be collected. Field samples need to be kept in shade and under cool temperature in field and transported to lab for processing within 24 h after collection. Each sample can be divided into two portions (A and B).

Portion A can be ground in liquid nitrogen and subject to whole DNA extraction. Traditional PCR can be performed prior to qPCR on the DNA extracts using one or more MLST primers. CS22.m2 primer pair is recommended since it amplifies both 3C genotype and non-3C genotype with similar efficiency and the amplicons show a size difference by genotype which is visible on agarose gel (Chapter 2). By using CS22.m2, detection of both genotypes in a single run is possible. Other MLST primer pairs than CS22.m2 can also be used for characterization of *C. flavescens* populations. Since there is no to little difference in amplicon sizes by genotype, digestion by restriction endonucleases targeting sequence differences between the two genotypes will be necessary to differentiate genotypes. Also there are several primer pairs (CS10.m1, CS22.m1 and EF1.m1) that were tested on both genotypes but not used for MLST analysis because they could not amplify non-3C type at all or not as efficiently as for 3C type (Chapter 2). Those can be used for selection of samples colonized by 3C type.

qPCR detection using pair h31.2 can be the next step to traditional PCR to minimize the number of reactions for the former since it is more expensive. qPCR can be performed only on the samples positive for MLST primers first. If the minimal amount detected was much higher than the theoretical detection limit of the qPCR, quantification
of samples negative for MLST primers by qPCR will be desired; if the minimal detection was around the theoretical detection limit, qPCR on negative samples for MLST will not be necessary.

Analysis on the banding patterns of traditional PCR (and restriction digestion if necessary) will show the presence or absence of *C. flavescens* populations of each genotype across the production system of wheat and barley.

qPCR quantification will provide data on abundance of 3C-type populations. Association of population type and abundance with types of crop, cultivar and tissue, timing of sample collection, geographic locations can also be investigated. The results will be informative for bioprospecting efforts in discovering novel *C. flavescens* biocontrol agents in terms of where and when to find them, how much of them can be found and the necessary sample sizes to successfully recover them. The question about how much native 3C-type populations will interfere with qPCR quantification of 3C arising from artificial inoculation can also be answered.

**Marker-assisted bioprospecting of *C. flavescens* isolates that are potential biocontrol agents**

**Marker-assisted isolation of *C. flavescens* isolates from wheat and barley followed by biocontrol efficacy test**

In the greenhouse bioassays described in Chapter 4, all the tested *C. flavescens* isolates strains showed biocontrol efficacy against Fusarium head blight (FHB), though there was high variation across runs of experiment. Those strains were isolated from diverse sources, including wheat anthers, dent corn kernels, soil and as a contaminant of an enzyme solution. This suggests that *C. flavescens* strains generally have biological control efficacy against FHB regardless of natural habitat from which it was isolated. In addition, novel strains within the same species may provide improved and/or complementary biocontrol efficacy than the identified strains. Therefore, bioprospecting
for more *C. flavescens* isolates from the environment may be a promising strategy for development of novel biopesticide against FHB. The MLST primers, which have specificity to *C. flavescens* over *C. neoformans*, can be used for marker-assisted bioprospecting of *C. flavescens* isolates. The primers are still likely to amplify other *Cryptococcus* species that are more closely related to *C. flavescens* than *C. neoformans* is. However, the possibility of those other species having biocontrol efficacy against FHB cannot be excluded. Therefore, the specificity of MLST primers should not be an issue for use in bioprospecting. Overall, I hypothesize that microbial recovery from wheat and possibly barley assisted by MLST primers will lead to isolation of novel *C. flavescens* biocontrol agents.

Bioprospecting of *C. flavescens* isolates can be conducted on Portion B of the samples collected for characterizing the diversity and abundance of natural *C. flavescens* populations in wheat and barley. If feasible, those portions can be frozen in 10% to 20% (v:v) glycerol suspensions at -80°C for ease of processing, since OH182.9 was actually isolated from glycerol stocks of wheat anthers. Isolation effort can be made only from the samples detected by qPCR as having greater than $10^4$ target gene copies per gram, or, if the number of such samples is very low, attempts can also be made on the samples showing lower levels of qPCR detection and/or those positive for non-3C genotype. Recovered isolates can be PCR-verified using MLST primers such as CS22.m2 and sequenced to confirm their identities of being a specific genotype of *C. flavescens*. Green house biocontrol efficacy test can be conducted to select for the isolates showing significant and consistent reduction of FHB damages.

Association of the numbers, identities and biocontrol efficacy of isolates with sample types, location and timing of sampling, and qPCR-detected population abundance can be analyzed. The percentage of selected isolates out of starting isolates at each screening stages and the frequency of confirmed isolates with satisfactory biocontrol efficacy versus total sampling efforts can be assessed. Whether the molecular markers will lead to non-*C. flavescens* isolates with biocontrol efficacy against FHB can also be discovered. These results will provide guidance to bioprospecting for biocontrol *C. flavescens* in
terms of how to cost-effectively minimize sampling efforts while maximize the possibility of recovering novel functional isolates.

**Develop genotypic markers more indicative of biocontrol efficacy of* C. flavescens *isolates**

The *C. flavescens* markers targeting the house-keeping regions are not highly discriminating for biocontrol efficacy (Figure 12, Chapter 4). The polymorphism in the genes actually affecting biocontrol activity is more likely to be indicative of efficacy. Such gene can be identified in the genome-wide gene discovery through transcriptomic and gene knockout, as described earlier in this chapter. The biocontrol-related loci can be sequenced in all the available *C. flavescens* strains (including the ones used in MLST analysis and those identified in the proposed diversity characterization research) to determine whether there is polymorphism. If there is, bioassay can be performed on the strains with different alleles to find out whether the polymorphism is correlated with biocontrol efficacy. The loci of good correlation can be the candidates for marker development. Primers targeting the polymorphic regions can be designed or restriction analysis can be designed to cut within the polymorphic regions. And those markers can be used to assist bio-prospecting, where only the isolates positive for the markers associated with high biocontrol activity can be recovered and subject to efficacy test. If funding is available, how the polymorphism contributes to differential biocontrol efficacy can be investigated by deciphering the protein structures and functions of the domains.

**Develop* C. flavescens*-selective media for bio-prospecting based on patterns of carbon utilization and chemical resistance shown in Biolog assay**

There is currently no selective media for *C. flavescens* that can eliminate large amounts of contaminants while recover most of the *C. flavescens* individuals from the environment. Such selective media will improve the efficiency of bio-prospecting for *C. flavescens* and can be developed based on the patterns of carbon utilization and chemical resistance shown in Biolog assay. The carbon sources well utilized by *C. flavescens* strains but not by other microorganisms commonly present on relevant isolation
substrates such as wheat heads can be candidates for ingredients of selective media. Similarly, the types of chemical stress that *C. flavescens* strains are resistant to while those other microorganisms are sensitive to can also be candidate ingredients. Different mixtures of candidate ingredients in different concentrations can be tested for recovering from pure *C. flavescens* and from relevant isolation substrates such as wheat heads, with comparison to general growth media. The ingredient combinations that give the optimal balance between minimizing contaminants and maximizing the diversity and abundance of *C. flavescens* can be developed into selective media. Such media can reduce the labor and resources needed for recovering novel *C. flavescens* isolates as potential biocontrol agents.

**Determining environmental fate of 3C-type biocontrol *C. flavescens* by monitoring their dispersal and survival in production system of wheat and barley**

The qPCR assay demonstrated meaningful dispersal pattern of 3C in field (Chapter 3). In combination of characterization of 3C type native population to improve interpretation, the qPCR assay can be used for environmental fate determination of 3C. This can complete part of the ecological risk assessment required by US Environmental Protection Agency to register 3C as a biopesticide.

In the study reported in Chapter 3 to test the validity of the qPCR assay in the field, a limited number of factors were used to minimize the complexity of experiment. This was to ensure that a meaningful pattern was detected by the qPCR assay can only be due to the quality of the assay itself. When applying this qPCR assay to determining environmental fate of 3C, conditions close to what growers will use in massive crop production are desired, such as irrigation and fungicide application. Artificial inoculation of FHB pathogen *Fusarium graminearum* can be included to test the correlation between 3C population level and disease damages. Various formulations of 3C inoculum and
different timing of inoculation can be used and their effects on 3C population dynamics in the field can also be tested to facilitate the development and registration of 3C as a biopesticide. Since barley is also a host of Fusarium head blight and 3C may have biocontrol efficacy on barley, quantification can be conducted on barley as well. I hypothesize that those aforementioned factors may affect the dispersal patterns of 3C throughout the crop production systems.

Wheat and barley fields can be established and sprayed with 3C in central areas and *F. graminearum* throughout the fields. Head and stem-leaf samples can be collected at a gradient of distances from the 3C-inoculated areas at a series of time points during the growing season. Sampling areas can be extended if 3C is detected at the farthest sampled distances from inoculated areas. After harvest, threshed grains and ground flours can be sampled. These products can be stored under typical conditions used by growers and/or industry, and sampled over time. Field residue and soil can be sampled at a series of time points postharvest until next growing season. Soil and water system in surrounding area can also be sampled after harvest, but sampling can be conducted before harvest as well if 3C is detected near the borders of fields. Samples can be stored at -80°C prior to homogenization and DNA extraction. Quantitative PCR using the 3C type-specific primer pair h31.2 can be performed on DNA extracts from the aforementioned samples. The abundance of native *C. flavescens* populations shown by the previously proposed characterization study on non-inoculated fields can be considered to properly interpret the contribution of artificial 3C inoculation to qPCR-detected populations.

Analysis on the levels of detected population across different distances to 3C-inoculated areas and the population dynamics over a series of time during the growing season will help understand the dispersal patterns of 3C under the conditions normally used in crop production and the effects of formulation and pathogen presence on 3C dispersal. Comparison of disease damages with qPCR-detected population level will serve as reference for deciding what level of 3C population is necessary to achieve optimal control of FHB. Population data on postharvest residue, soil and surrounding water system will inform how well 3C survives in the environment after the growing season and how much of 3C spread beyond wheat-grown area. Monitoring 3C population
changes on threshed grains and ground flours will answer the question how long 3C can persist on these end products and whether it should be a concern of food safety. This study will provide information on exposure levels of the environment to 3C and help format ecological risk assessment as part of EPA requirement for biopesticide registration (http://www.epa.gov/pesticides/ecosystem/ecorisk.htm).

Breeding of *C. flavescens* for higher biocontrol efficacy

Mating test revealed two inter-compatible and self-incompatible mating types in *C. flavescens* (Kurtzman, 1973). Multilocus sequence typing showed that among all the strains of which the genotype and mating type were both known, both the strains of 3C genotype were mating type a while all the strains of non-3C genotype were mating type α (Chapter 4). The genetic divergence indicated that there was no to little sexual recombination between the two mating types in the environment without human intervening. Heterosis has been widely used in plant breeding to create hybrid that inherit the advantageous traits from both parents and thus perform better than the parents. Whether crossing compatible *C. flavescens* strains will result in hybrids with better biocontrol performance than the parents can be investigated. Understanding the mating of *C. flavescens* will help to breed *C. flavescens* hybrids.

Identify mating loci and develop sequence typing methods to determine mating types

Though mating types can be identified through mating test, whether two *C. flavescens* strains produce mating responses largely depends on test conditions (Kurtzman, 1973). This means that if two strains do not produce mating responses under a certain condition, either other conditions need to be tested or the risk of false negative identification needs to be taken. Moreover, mating test for *C. flavescens* typically takes 2 to 6 weeks (Kurtzman, 1973). In contrast, identification of mating types by mating locus is more reproducible regardless of conditions and less time-consuming.
To identify the mating loci in *C. flavescens*, the mating types of the more available *C. flavescens* strains than those studied in (Kurtzman, 1973) should to be determined, especially 3C since the genomic sequence of 3C is available. This can be accomplished through mating test if conditions inductive to mating responses of those strains can be found. This may take some investigation since the conditions that were inductive in 1973 may not still be inductive now after the strains have been subcultured and stored frozen over the years (Cletus Kurtzman, personal communication).

The genetic composition of *C. flavescens* mating locus has not been investigated. But the mating locus of *C. neoformans* has been well studied and shown to span approximately 105 to 130 kb, including 24 to 32 genes (Lengeler et al., 2002).

Discontiguous megablast and tblastx search was performed by querying the mating locus and flanking regions of *C. neoformans* JEC20 (bases 1,514,119 to 1,652,717 on the NCBI entry of accession number AE017344) against *C. flavescens* 3C genome. Most of the hits were throughout the whole length of scaffold CAUG01000348 (NODE_681_length_189955_cov_46.433445 in original assembly) but in low synteny (data not shown). This suggests that the mating locus of *C. flavescens* may be located on the chromosome where scaffold CAUG01000348 belongs and that there had been extensive rearrangements within this region between the two species. Putative functions of the genes within this region of 3C can be predicted through blastx and/or tblastx search against NCBI non-redundant protein sequences (nr) and/or nucleotide collection (nr/nt) databases, respectively. For the 3C genes that are homologous to those that belong to mating loci of other organisms, gene knockout analysis can be performed (once the mating type of 3C is determined) and determine whether those genes affect mating of 3C to its opposite mating type. The genes of which knockout results in abortion of mating while complementation restores mating should be parts of the mating locus.

“Yeast breeding” – Does hybrid progeny provide better biocontrol against FHB?

Though the *C. flavescens* strains currently available in NRRL collection did not show significant difference in biocontrol efficacy against FHB (Chapter 4), other strains that will be isolated during the diversity characterization work may show superior
performance. The strains with higher efficacy and/or other desired traits such as fungicide tolerance can be selected and subject to mating with compatible mating types. Progeny can be isolated as single colonies and tested in bioassays to check whether they show more desired phenotypes, especially higher biocontrol efficacy, than parents. If heterosis is observed in *C. flavescens*, crossing can be used to breed novel *C. flavescens* hybrids that are potentially better active ingredients of biopesticides than natural isolates.

**Conclusions**

The overarching goal for the whole research on *C. flavescens* is to facilitate its development and registration as a biopesticide. Genome-wide discovery of genes related to biocontrol of 3C will provide insight into biocontrol mechanism of *C. flavescens* against Fusarium head blight. This, in turn, will enable the improvement in formulation and application of *C. flavescens*-based biopesticide according to the mode of action of the active ingredient. Characterizing the diversity and abundance of natural *C. flavescens* population will serve as the basis for discovery of novel *C. flavescens* strains that are potentially biocontrol agents. Crossing *C. flavescens* strains of compatible mating types may give rise to hybrids with superior efficacy than the parents. These novel isolates and hybrids can then be used to develop new biopesticides with improved and/or complementary efficacy compared to existing ones. In addition, determining environmental fate of 3C throughout the wheat production system will facilitate biopesticide registration of 3C with EPA, which enables commercialization of *C. flavescens*-based biopesticide.
References


52. Lodder J, Kreger-van Rij NJW (eds.), 1952. The Yeasts, a Taxonomic Study. , Amsterdam, Netherland.


87. **Schisler DA, Paul, Boehm, Bradley, Dunlap.** 2010. Colonization of wheat heads by antagonist *Cryptococcus flavescens* OH 182.9 when applied alone or in combination with different concentrations of Prosaro® and the effect on Fusarium head blight development in field-grown wheat, p. 98-102. In Anonymous , Milwaukee, WI.


Appendix A: computer scripts used in genomic analysis

Python script for converting multiple lines of sequences between two headers into one line

In a standard .fasta file, each sequence entry is one header line followed by one sequence line without returns. Sometimes assembly algorithms such as Velvet will produce .fasta files where one header line is followed by multiple sequence lines (separated by returns) that are parts of a continuous sequence. This violation of standard may confuse .fasta parsing for downstream processing and analysis, including the two scripts below. Here is a python (.py) script (in Arial font) for converting the multiple lines into one line written for the assembly where header lines start with “>NODE_xxx_” (“xxx” is scaffold identifier):

# Content between two @s is to be determined by you or your FASTA file. #

# Load assembly file to Python program and assign it to the variable "MultiLineAssembly" #
MultiLineAssembly = open('@ the .fasta assembly you want to convert @')

# Assign the "one line per contig/scaffold" .fa file you want to write in (@identifier@.fa) to the variable OneLine #
OneLine = open('Cf_31_300r.fa','r+')

# This is to start writing from the end of the file. Include this line if you don't want to over-write anything already present in the file #
OneLine.read()

# Define the vector string variable
ToWrite = "

# Start read line by line of your assembly and to write in your "one line per contig/scaffold" .fa file #
for StringLine in MultiLineAssembly:
    BaseEnd = len(StringLine) - 1
    if StringLine[0] == '>':
        WriteLine = '\n' + StringLine
    else:
        WriteLine = StringLine[0:BaseEnd]
    ToWrite = ToWrite + WriteLine

# Get rid of the first line return #
ToWrite = ToWrite[1:]
# Put in the last nucleotide if there is no line return at the end of the multi-line assembly file #
if StringLine[len(StringLine) - 1] != '\n':
    ToWrite = ToWrite + StringLine[len(StringLine) - 1]
else:
    pass
OneLine.write(ToWrite)

# Seek back to the beginning of one-line.fa #
OneLine.seek(0)
# Read the whole one-line.fa to finalize writing #
OneLine.read()

Python script for splitting scaffolds at > 10 consecutive ‘N’s

When Velvet with Pebble and Rock Band assembles paired-end reads from high-throughput sequencing, it will connect contigs using the spatial relationship of read pairs into scaffolds and fill the gaps with stretches of ≥ 10 consecutive character ‘N’s where no sequence information is available (Zerbino 2009). European Molecular Biology Library (EMBL) requires genome sequences free of stretches > 10 consecutive ‘N’s to be
submitted, therefore, scaffolds have to be split into contigs at > 10 consecutive ‘N’s. Here is a python (.py) script (in Arial font) for splitting scaffolds written for the assembly where header lines start with “>NODE_@xxx_” (“xxx” is scaffold identifier):

# Content between two @s is to be determined by you or your FASTA file. #

# Load assembly file to Python program and assign it to the variable "UnsplitAssembly" #
UnsplitAssembly = open('@ the .fasta assembly you want to split @')

# Assign the _Split10plusN.fa file you want to write in (@identifier@_Split10plusN.fa) to the variable Split10plusN #
Split10plusN = open('@ the .fasta you want to store your split assembly in @','r+')

# This is to start writing from the end of the file. Include this line if you don't want to over-write anything already present in the file #
Split10plusN.read()

# Start read line by line of your assembly and to write in your "@identifier@_Split10plusN.fa" file #
for StringLine in UnsplitAssembly:
    # Extract "NODE_@number@" (variable "unsplitNodeName" in script) as the root for "splitNodeName" #
    if StringLine[0] == '>':
        unsplitNodeNameEnd = 6
        while StringLine[unsplitNodeNameEnd] != '_':
            unsplitNodeNameEnd = unsplitNodeNameEnd + 1
        unsplitNodeName = 'NODE_' + StringLine[6:unsplitNodeNameEnd]
    # Move onto the sequence line next to the "NODE_@number@" line #
    else:
        # Define/return the values in the pointer variables as/to starting stage #
        splitNode_beg, splitNode_end, splitNode_id = 0,0,1
        # To read from the start to the end of this line #
        while splitNode_end < len(StringLine):
            # Read and summarize the component (determined sequence, non-
            "NNN...") #
            if StringLine[splitNode_end] != 'N':
splitNode_end = splitNode_end + 1

# Reach gap ("NNN...")#
else:
    # Start calculating gap length#
    gap_beg = gap_end = splitNode_end
    while StringLine[gap_end] == 'N':
        gap_end = gap_end + 1
    # If gap length is greater than 10, split a "split Node" out of the
"split Nodes" before this gap#
    if gap_end - gap_beg > 10:
        # Write the "splitNodeName" and the sequence of the
"split Node" to _Split10plusN.fa#
        ToWrite = '>' + unsplitNodeName + '-' + str(splitNode_id) + '
        + '
        + StringLine[splitNode_beg:splitNode_end] + 'n'
        Split10plusN.write(ToWrite)
        # Get splitNode_beg and splitNode_id ready for the next
'split Node'#
        splitNode_beg = gap_end
        splitNode_id = splitNode_id + 1
    # If gap length is 10 or less, do nothing#
else:
    pass
    # Get splitNode_end ready for the next step#
    splitNode_end = gap_end
    # Write the last "split Node" in the respective line (scaffold), if the "unsplit Node"
was ever split, or the whole "unsplit Node", if it has never been split, to _Split10plusN.fa#
if splitNode_beg != 0:
    ToWrite = '>' + unsplitNodeName + '-' + str(splitNode_id) + '
    + '
    + StringLine[splitNode_beg:splitNode_end]
else:
    ToWrite = '>' + unsplitNodeName + '

StringLine[splitNode_beg:splitNode_end]
Split10plusN.write(ToWrite)

# Seek back to the beginning of _Split10plusN.fa#
Split10plusN.seek(0)
# Read the whole _Split10plusN.fa to finalize writing #
Split10plusN.read()

## Python script for making .agp file for your assembly

EMBL requires a .agp file describing the genome segments (can be chromosome, scaffolds, contigs) to be submitted together with genome sequences. Here is a python (.py) script (in Arial font) for splitting scaffolds written for the assembly where header lines start with “>NODE_ xxx_” (“xxx” is scaffold identifier):

# Content between two @s is to be determined by you or your FASTA file. #

# Load assembly file to Python program and assign it to the variable "Assembly" #
Assembly=open('@ the .fasta assembly you want to write .agp for @')

# Assign the .agp file you want to write in (@identifier@.agp) to the variable forAGP #
forAGP = open('@ the destination .agp file you want to write in @','r+')

# This is to start writing from the end of the file. Include this line if you don't want to over-write anything already present in the file #
forAGP.read()

# Start read line by line of your assembly and to write in your "@identifier@.agp" file #
for StringLine in Assembly:
    # Extract "NODE_@number@" as the identifier for "object" (variable "objectNodeName" in script) #
    if StringLine[0] == '>':
        objectNodeNameEnd = 6
        while StringLine[objectNodeNameEnd] != '_':
            objectNodeNameEnd = objectNodeNameEnd + 1
        objectNodeName = 'NODE_' + StringLine[6:objectNodeNameEnd]
    # Move onto the sequence line next to the "NODE_@number@" line #
    else:
        # Define/return the values in the pointer variables as/to starting stage #
part_number, objectCMPNT_beg, objectCMPNT_end, component_id = 1,1,0,1
# To read from the start to the end of this line #
while objectCMPNT_end < len(StringLine):
    # Read and summarize the component (determined sequence, non-
"NNN...") #
    if StringLine[objectCMPNT_end] != 'N':
        objectCMPNT_end = objectCMPNT_end + 1
    # Reach gap ("NNN...") #
    else:
        # Write the summary of the above component to .agp #
        ToWrite = objectNodeName + 't' + str(objectCMPNT_beg) + 't' + str(objectCMPNT_end) + 't' + str(part_number) + 'tW\t' + objectNodeName+'.' + str(component_id) + 't1t' + str(objectCMPNT_end - objectCMPNT_beg + 1) + 'tn'
        forAGP.write(ToWrite)
        # Get the numeral-column values ready for the following gap #
        part_number = part_number + 1
        objectCMPNT_beg = objectCMPNT_end + 1
        # Get component_id ready for next component#
        component_id = component_id + 1
        # Continue to read the line, staring with an "N" #
        while StringLine[objectCMPNT_end] == 'N':
            objectCMPNT_end = objectCMPNT_end + 1
            # Reach a non-"N" character, start of next component #
            # Write the summary of the above gap to .agp #
            ToWrite = objectNodeName + 't' + str(objectCMPNT_beg) + 't' + str(objectCMPNT_end) + 't' + str(part_number) + 'tN\t' + str(objectCMPNT_end - objectCMPNT_beg + 1) + 'tscaffold\tyes\tpaired-ends\n'
            forAGP.write(ToWrite)
            # Get the numeral-column values ready for the following
            component #
            part_number = part_number + 1
            objectCMPNT_beg = objectCMPNT_end + 1
        # Adjust the objectCMPNT_end value if it is miscounted due to the \n' at the end of each line #
        if StringLine[objectCMPNT_end - 1] == '\n':

128
objectCMPNT_end = objectCMPNT_end - 1
else:
    pass

# Write the last component in the respective line (scaffold) to .agp #
ToWrite = objectNodeName + '\t' + str(objectCMPNT_beg) + '\t' + str(objectCMPNT_end) + '\t' + str(part_number) + '\tW\t' + objectNodeName + '.' + str(component_id) + '\t1\t' + str(objectCMPNT_end - objectCMPNT_beg + 1) + '\t+n'
    forAGP.write(ToWrite)

# Seek back to the beginning of .agp #
forAGP.seek(0)
# Read the whole .agp to finalize writing #
forAGP.read()

Sample .agp files for scaffolds (assembly with > 10 consecutive ‘N’)

... NODE_3033 1 22741 1 W NODE_3033.1 1 22741 +
NODE_3033 22742 22751 2 N 10 scaffold yes paired-ends
NODE_3033 22752 28575 3 W NODE_3033.2 1 5824 +
NODE_3033 28576 28585 4 N 10 scaffold yes paired-ends
NODE_3033 28586 84629 5 W NODE_3033.3 1 56044 +
NODE_3033 84630 84640 6 N 11 scaffold yes paired-ends
NODE_3033 84641 126308 7 W NODE_3033.4 1 41668 +
NODE_3033 126309 126318 8 N 10 scaffold yes paired-ends
NODE_3033 126319 126383 9 W NODE_3033.5 1 65 +
NODE_3100 1 2500 1 W NODE_3100.1 1 2500 +
NODE_3100 2501 2510 2 N 10 scaffold yes paired-ends
NODE_3100 2511 2812 3 W NODE_3100.2 1 302 +
...

Sample .agp files for contigs (assembly without > 10 consecutive ‘N’)
... NODE_3033-1 1 22741 1 W NODE_3033-1.1 1 22741 +
NODE_3033-1 22742 22751 2 N 10 scaffold yes paired-ends
NODE_3033-1 22752 28575 3 W NODE_3033-1.2 1 5824 +
NODE_3033-1 28576 28585 4 N 10 scaffold yes paired-ends
NODE_3033-1 28586 84629 5 W NODE_3033-1.3 1 56044 +
NODE_3033-2 1 41668 1 W NODE_3033-2.1 1 41668 +
NODE_3033-2 41669 41678 2 N 10 scaffold yes paired-ends
NODE_3033-2 41679 41743 3 W NODE_3033-2.2 1 65 +
NODE_3100 1 2500 1 W NODE_3100.1 1 2500 +
NODE_3100 2501 2510 2 N 10 scaffold yes paired-ends
NODE_3100 2511 2812 3 W NODE_3100.2 1 302 +
...

Customization of existing Python scripts for converting BLAST .xml into tab-separated values (also known as .tsv, .tab or .txt)

.xml format of BLAST search results include the most complete set of information among all the formats available for downloading from NCBI website. BLAST .xml resulting from querying uncharacterized genome against public database is useful for genome annotation. Parsing BLAST .xml and converting it into other formats easier to manipulate, such as tab-separated values (also known as .tsv, .tab or .txt), is usually desirable for high-throughput annotation. A BLAST .xml converter (Copyright 2012 by Wibowo Arindrarto) is available as part of Biopython 1.61 package (http://biopython.org/wiki/Download). Specifically, the portions of the converter used in this thesis were 'blast_xml.py' and 'blast_tab.py' under ‘Lib\site-packages\Bio\SearchIO\BlastIO’ within the directory where Python 3.3.2 is installed (www.python.org). The .tsv files converted from .xml by the original Biopython scripts does not include annotation of hit (‘Hit_def’ field in .xml), thus cannot be conveniently
used for genome annotation. Therefore, the original scripts were customized to include annotation of hit in resulting .tsv. Script content is shown in Arial font.

**Customization of ‘blast_xml.py’**

*Change 1*

**In original script:**
Line 14, no content

**In customized script:**
Line 14-15:
# ^_^ To show in IDLE (Python GUI) that the module is customized by Xiaoqing Rong
print('blast_xml was customized by Xiaoqing Rong')
Line 16-17, no content

Starting from Line 15 in original and Line 18 in customized, Line \( x \) in original is the same as Line \( x+3 \) in customized.

*Change 2*

**In original script:**
Line 50:
#’Hit_def’: (‘description’, str), # not set by this dict

**In customized script:**
Line 53-54:
# ^_^ Blockage on parsing ‘Hit_def’ was released by removing “#”.
’Hit_def’: (‘description’, str), # ^_^ originally “not set by this dict”

Starting from Line 51 in original and Line 55 in customized, Line \( x \) in original is the same as Line \( x+4 \) in customized.
Change 3

In original script:
Line 374-377:
    try:
        hit_desc = id_desc[1]
    except IndexError:
        hit_desc = ''

In customized script:
Line 378-382:
    # ^_^ suppress the following by adding "# ^o^ " to start
    # ^o^ try:
        # ^o^ hit_desc = id_desc[1]
    # ^o^ except IndexError:
        # ^o^ hit_desc = ''

Starting from Line 378 in original and Line 383 in customized, Line $x$ in original is the same as Line $x+5$ in customized.

Customization of ‘blast_tab.py’

Change 1

In original script:
Line 14, no content

In customized script:
Line 14-15:
    # ^_^ To show in IDLE (Python GUI) that the module is customized by Xiaoqing Rong
    print('blast_tab was customized by Xiaoqing Rong')
Line 16-17, no content
Starting from Line 15 in original and Line 18 in customized, Line \( x \) in original is the same as Line \( x+3 \) in customized.

\textit{Change 2}

\textbf{In original script:}

Line 38:

\begin{verbatim}
'mismatches': 'mismatch',
\end{verbatim}

\textbf{In customized script:}

Line 41-44:

\begin{verbatim}
# ^_^ add 'sseqdescrp' for getting Hit_def
'query/sbjct frames': 'sseqdescrp',
# ^_^ end of change
'mismatches': 'mismatch',
\end{verbatim}

Starting from Line 38 in original and Line 44 in customized, Line \( x \) in original is the same as Line \( x+6 \) in customized.

\textit{Change 3}

\textbf{In original script:}

Line 68:

\begin{verbatim}
'sacc': ('accession', str),
\end{verbatim}

\textbf{In customized script:}

Line 74-77:

\begin{verbatim}
# ^_^ add 'sseqdescrp' for getting Hit_def
'sseqdescrp': ('description', str),
# ^_^ end of change
'sacc': ('accession', str),
\end{verbatim}
Starting from Line 68 in original and Line 77 in customized, Line $x$ in original is the same as Line $x+9$ in customized.

**Change 4**

**In original script:**

Line 104-105:

```python
_DEFAULT_FIELDS = ['qseqid', 'sseqid', 'pident', 'length', 'mismatch','gapopen', 'qstart', 'qend', 'sstart', 'send', 'evalue', 'bitscore']
```

**In customized script:**

Line 113-115:

```python
# ^_^ 'sseqdescrp' was added to _DEFAULT_FIELDS
_DEFAULT_FIELDS = ['qseqid', 'sseqid', 'sseqdescrp', 'pident', 'length', 'mismatch','gapopen', 'qstart', 'qend', 'sstart', 'send', 'evalue', 'bitscore']
```

Starting from Line 105 in original and Line 115 in customized, Line $x$ in original is the same as Line $x+10$ in customized.