Copyright © 2008 by Jo Anne Crouch
All rights reserved

Evolution of *Colletotrichum* species inhabiting grasses in diverse ecosystems

by

Jo Anne Crouch

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

written under the direction of

Bradley I. Hillman and Bruce B. Clarke

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2008

Abstract of the dissertation

Evolution of *Colletotrichum* species inhabiting grasses in diverse ecosystems

By Jo Anne Crouch

Dissertation Directors: Bradley I. Hillman and Bruce B. Clarke

Fungi in the genus *Colletotrichum* are destructive pathogens, afflicting over 900 plant species worldwide, including numerous grasses and cereal crops. In this dissertation I reconstruct the evolution of grass-inhabiting *Colletotrichum* and relate it to pathogenicity. I have undertaken this research to (1) understand how graminicolous *Colletotrichum* were shaped by evolutionary processes such as population divergence, gene flow, mutation, recombination and speciation; and (2) evaluate whether the lifestyles of these fungi can be correlated with genotypic or genomic signatures, life history or ecological adaptations. These broad objectives were centered on the application-based model of *C. cereale*, a recently emerged pathogen of the golf course turfgrasses. Novel tools, including sequence-based markers from four protein coding genes and five transposon species, transposon RFLPs, and microsatellite markers were developed and deployed. Although ITS

sequences are currently the most common method of classifying *Colletrichum* species, this research highlighted the potential inaccuracy of ITS-based classification. The unreliability of 47% of *Colletotrichum* ITS sequences from public databases suggested a proliferation of compromised species identifications. Comparison with multilocus phylogenies showed *Colletotrichum* ITS data are insufficient for the task of taxonomic resolution. In addition, these studies demonstrated traditional classification tools (*i.e.*, morphology and host range) are subject to convergent evolution. Phylogenetic reconstructions showed that graminicolous *Colletotrichum* underwent a prominent historical split, separating cool-season grass-associated taxa from lineages inhabiting warm-season grasses. Eight novel *Colletotrichum* species were identified and described; two species were emended. *C. cereale* populations were found in native grasses, cereal crops and turfgrass environments, but disease was limited to turfgrass. Genotypic data, along with the detection of the meiosis-specific repeat-induced point mutation process provided evidence of recombination in *C. cereale*, a fungus long presumed asexual. Extreme differentiation between locally-adapted populations indicated that asymptomatic grasses are unlikely reservoirs of infectious disease propagules that could initiate disease in turf. But gene flow from the generalist *C. cereale* founder population and specialized genotypes provides an indirect pathway for genetic exchange between otherwise isolated populations. Together, these studies contribute substantially to the growing number of genomic resources available for this increasingly important evolutionary research system.

Acknowledgments

Chapter One: This work was published as: Crouch, J.A., Clarke, B.B., and Hillman, B.I. (2006) Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and maize. Phytopathology: 96(1):46-60 (Appendix 2). For this study, I am grateful to Lisa Vaillancourt for her suggestion to use the *C.graminicola* mating-type locus as a phylogenetic marker in this analysis and for sharing a pre-publication version of reference 25 while the manuscript was in review. I thank Richard Buckley and Sabrina Tirpak of the Rutgers Disease Diagnostic Clinic for identifying and supplying *Colletotrichum*-infected turfgrass specimens, and to several individuals who generously provided fungal cultures for use in this study: Jim Baird, Gary Bergstrom, Tom Hsiang, Noel Jackson, Randy Kane, Peter Oudemans, Wakar Uddin, Lisa Vaillancourt, and Frank Wong. Special thanks to the herbarium staff of the New York Botanical Garden for their assistance with herbarium specimens. We also thank Karen Garrett and Ned Tisserat for the use of laboratory facilities and assistance with sampling conducted at the Konza Prairie Biological Station (a preserve of The Nature Conservancy managed by the Division of Biology, Kansas State University), and Karl Kjer for assistance with the data sets, phylogenetic analyses, and for comments on the manuscript. I also thank Tom Harrington, Jim White, and three anonymous reviewers, who provided valuable discussions, comments, and criticisms that greatly improved the quality of this manuscript, and my co-authors Bruce B. Clarke and Bradley I. Hillman. Thanks also to Michael Milgroom, who, as senior editor for *Phytopathology*, helped develop the first draft of the manuscript into a polished and comprehensive paper.

Chapter Two: This work was published as: Crouch, J.A., Glasheen, B.M., Giunta, M.A., Clarke, B.B. and Hillman, B.I. (2008) The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. Fungal Genetics and Biology 45:190–206 (Appendix 3). For this study, I thank Lisa Vaillancourt for providing the *C. graminicola* and *C. sublineolum* cultures used in this study, and my co-authors Bernadette M. Glasheen, Michael A. Guinta, Bruce B. Clarke and Bradley I. Hillman. Thanks also go to Michael Milgroom for his help in reorganizing an earlier, lengthier version of this manuscript and to Linda Kohn, who, as editor for Fungal Genetics and Biology, helped develop the final published version of the manuscript.

Chapter Three: This work is currently in press, and will be published as: Crouch, J.A., Glasheen, B.M., Uddin, W., Clarke, B.B., and Hillman, B.I. (2008) Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles. In press, Crop Science (Appendix 4). For this research, I thank Lisa Vaillancourt for providing the *C. graminicola* and *C. sublineolum* cultures; the National Institute of Agrobiological Sciences Genebank of Ibaraki, Japan for the *C. falcatum* cultures used in this study and my co-authors Bernadette M. Glasheen, Wakar Udin, Bruce B. Clarke and Bradley I. Hillman.

Chapter Four: This work has been submitted for review to the journal *Molecular Ecology*: Crouch, J.A., Tredway, L.P., Clarke, B.B. and Hillman, B.I. Ecological specialization drives the evolution of the pathogenic fungus *Colletotrichum cereale* and allied taxa across

diverse grass communities. For this study, I thank Gary Bergstrom, Tom Hsiang, Noel Jackson, John Kaminski, Randy Kane, Peter Oudemans, Dave TeBeest, Maria-Tomaso-Peterson, Wakar Uddin, Lisa Vaillancourt, and Frank Wong for sharing many of the fungal cultures used in this study; Aaron Newland, Joseph Roberts, and Anjali Majumdar for technical assistance; Josh Honig for running the sequence reactions; and Rich Buckley and Sabrina Tirpak of the Rutgers Plant Diagnostic Clinic for supplying *Colletotrichum*-infected turfgrass specimens. For allowing us to sample their sites, I thank the Konza Prairie Biological Station, the Indian Boundary Prairie, the Nachusa Grasslands, the Willa Cather Memorial Prairie (preserves of The Nature Conservancy in Manhattan, KS; Chicago, IL; Franklin Grove, IL; and Red Cloud, NE respectively), the Peach Point Wildlife Management Area (Freeport, TX), the Wallkill Wildlife Management Area (Sussex Co., NJ), and The Land Institute (Salina, KS). I also thank my co-authors, Lane Tredway, Bruce Clarke and Bradley Hillman for their contributions to this work.

Chapter Five: Thanks to Josh Honig for several helpful discussions about microsatellite markers and suggestions for multiplexing strategies and the incorporation of M13 sequences in the primer sequences for fluorescent dye labeling.

Chapter Six: Thanks to Jim White for helpful discussions, Patricia Eckel of the Missouri Botanical Garden for the Latin translations, Josh Honig for running the sequence reactions, and the New York Botanical Garden Steere Herbarium and the USDA National Fungus Collection (BPI) for the loan of herbarium specimens.

Chapter Seven: Thanks to Peter Johnston, Gary Samuels and Amy Rossman for ongoing discussions about *Colletotrichum* systematics, and to Peter Oudemans for introducing me to the world of *Plant Disease* notes.

This work was funded by grants from the Rutgers Center for Turfgrass Science to Bradley I. Hillman and Bruce B. Clarke and by the New Jersey Agricultural Experiment Station. I gratefully acknowledge financial support during the course of my graduate studies provided by a U.S. Environmental Protection Agency Science to Achieve Results (STAR) Graduate Fellowship, a Rutgers Excellence Fellowship, the Robert White-Stevens Fellowship and a Land Institute Natural Systems Agriculture Graduate Fellowship.

Table of contents

Abstract of the dissertation ii

Acknowledgments iv

Table of contents viii

List of tables xviii

List of figures xxi

Preface: Introduction to the dissertation: Evolution of *Colletotrichum* species inhabiting grasses in diverse ecosystems

P.1 Introduction 1

P.2 Thesis statement and objectives 1

P.3 The genus *Colletotrichum*: Pathogens causing disease worldwide 3

P.4 Taxonomy, classification and species concepts in the genus *Colletotrichum* 4

P.5 References 5

Chapter 1: Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn

ABSTRACT 7

1.1 Introduction 8

1.2 Materials and methods

1.2.1 Taxon sampling for molecular analysis 11

1.2.2 Outgroup taxa 12

1.2.3 DNA isolation, amplification and sequencing..... 13

1.2.4 Phylogenetic analyses..... 15

1.2.5 Estimating recombination and its impact on phylogenetic inference 16

1.2.6 Morphological analysis 18

1.3 Results

1.3.1 Dataset characteristics and analysis..... 19

1.3.2 Phylogenetic relationships..... 23

1.3.3 Reticulation and incompatibility in the dataset 24

1.4 Taxonomy

1.4.1 *Colletotrichum cereale* Manns *sensu lato* Crouch, BB Clarke and Hillman..... 26

Emended description 27

Type specimen..... 28

Comments 29

1.4.2 *Colletotrichum graminicola* (Ces.) G.W. Wilson *sensu stricto* Sutton 30

Emended description 30

Epitype establishment 31

Comments 31

1.4.3 *Colletotrichum sublineolum* Henn *apud.* Kabat & Bub..... 32

Emended description 32

Epitype establishment 33

Comments 33

1.5 Discussion

2.5.1 Species boundaries 34

2.5.2 Unresolved, potentially species-level divergencse 35

2.5.3 The evolutionary history of *C. cereale* 38

1.6 References 41

Chapter 2: The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen

ABSTRACT 59

2.1 Introduction 59

2.2 Materials and methods

2.2.1 Construction of genomic DNA libraries and the identification of repetitive transposon DNA 63

2.2.2 Sequence analysis 64

2.2.3 Estimates of repeat-induced point mutation 65

2.2.3 Estimates of incongruent evolutionary relationships caused by homoplasy or recombination 67

3.2.4 Nucleotide sequence assession numbers 68

2.3 Results

2.3.1 Identification and nomenclature of TEs from the *C. cereale* and *C. sublineolum* genomes 68
2.3.2 Degenerate transposons are identified from the *C. cereale* clade B genome 70
2.3.3 Intact transposon sequences are recovered from the *C. cereale* clade A genome 71
2.3.4 Assessment of RIP-like patterns in the transposon sequences 73
2.3.5 Genomic population analysis using *Ccret2*$^{POL2/3}$ sequence data 75
2.4 Discussion 77
2.5 References 83

Chapter 3: Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles

ABSTRACT 105
3.1 Introduction 105
3.2 Materials and methods
3.2.1 Fungal cultures 108
3.2.2 Restriction fragment length polymorphism (RFLP) analyses 108
3.2.3 Phylogenetic analyses 109
3.2.4 Nucleotide sequences 111
3.3 Results
3.3.1 Phylogenetic assessment of populations using sequence data 111

3.3.2 Limited distribution of the TE sequences *Collect*^{I29} and *Ccret2*^{DBP16} 112

3.3.3 The retrotransposons *Ccret1*^{DBP6} and *Ccret2*^{A15} are found in both *C. cereale* lineages 113

3.3.4 Estimates of population subdivision using the retrotransposon RFLP datasets 114

3.4 Discussion 115

3.5 References 118

Chapter 4: Ecological specialization drives the evolution of the pathogenic fungus *Colletotrichum cereale* and allied taxa across diverse grass communities

ABSTRACT 126

4.1 Introduction 127

4.2 Materials and methods

4.2.1 Sample collection 132

4.2.2 Molecular data 133

4.2.3 Phylogenetic analyses 134

4.2.4 Population analyses 136

4.3 Results

4.3.1 Species boundaries and the evolution of grass-associated *Colletotrichum* 137

4.3.2 Genetic differentiation among *C. cereale* populations 140

4.4 Discussion and conclusions

4.4.1 The impact of environmental context and host specialization on the evolution of grass-inhabiting *Colletotrichum* 145

4.4.2 The origin of *C. cereale* populations in diverse ecosystems 146

4.4.3 The influence of alternative *C. cereale* ecotypes on the initiation and maintenance of anthracnose disease in turfgrass and other ecosystems. 147

4.5 References 150

Chapter 5: The development of microsatellite markers as tools to expand population analyses of *Colletotrichum cereale* in turfgrass, cereal crops and natural ecosystems

ABSTRACT 177

5.1 Introduction 178

5.2 Materials and methods 179

5.3 Results and discussion 182

5.4 References 187

Chapter 6 Systematic analysis of the falcate-spored graminicolous *Colletotrichum* and a description of five new species from warm season grasses

ABSTRACT 204

6.1 Introduction 205

6.2 Materials and methods

6.2.1 Fungal isolates 206

6.2.2 Molecular phylogenetic analysis 207

6.2.3 Morphological analysis 208

6.3 Results and discussion 209

6.4 Taxonomy 213

6.4.1 *Colletotrichum hanaui* Crouch, JF White, BB Clarke and Hillman *sp. nov.* 215

6.4.2 *Colletotrichum nicholsonii* Crouch, JF White, BB Clarke and Hillman *sp. nov* 217

6.4.3 *Colletotrichum jacksonii* Crouch, JF White, BB Clarke and Hillman sp. nov. 219

6.4.4 *Colletotrichum miscanthi* Crouch, JF White, BB Clarke and Hillman *sp. nov* 221

6.4.5 *Colletotrichum axonopodi* Crouch, JF White, BB Clarke and Hillman *sp. nov* 222

6.4.6 *Colletotrichum elesines* Pavgi & UP Singh, emend. Crouch, JF White,

BB Clarke and Hillman 224

6.5 References 225

Chapter 7 Can species of *Colletotrichum* be uniquely identified through ITS sequence data? A case study using the falcate-spored graminicolous *Colletotrichum* group

7.1 Introduction 272

7.2 Materials and methods

7.2.1 Fungal isolates 277

7.2.2 Molecular phylogenetic analysis 277

7.2.3 Evaluation of ITS sequences curated by NCBI GenBank 278

7.3 Results

7.3.1 The ITS gene genealogy as a phylogenetic indicator of species boundaries and supraspecific groups 279

7.3.2 Database searches using ITS sequences from authentic FG *Colletotrichum* species 282

7.4 Discussion and conclusions 290

7.5 References 290

Appendices

Appendix 1: GenBank accession numbers of nucleotide sequence data generated by the dissertation research 304

Appendix 2: Publication version of Chapter 1:

Crouch, J.A., Clarke, B.B., and Hillman, B.I. (2006) Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass andcorn. Phytopathology: 96(1):46-60 305

Appendix 3: Publication version of Chapter 2:

Crouch, J.A., Glasheen, B.M., Giunta, M.A., Clarke, B.B. and Hillman, B.I. (2008) The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. Fungal Genetics and Biology 45:190–206. 320

Appendix 4: Publication version of Chapter 3:

Crouch, J.A., Glasheen, B.M., Uddin, W., Clarke, B.B., and Hillman, B.I. (2008) Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles. In press with Crop Science 337

Curriculum vitae 345

List of tables

Table	Title	Page #
Chapter 1		
1.1	Sources of *Colletotrichum* isolates used in this study	45
1.2	Characteristics of the genomic regions used in this investigation	49
1.3	Best fit evolutionary models used in Bayesian analysis as determined by the Akaike information criterion (AIC) in Modeltest	50
Chapter 2		
2.1	Primer sequences used in this study	89
2.2	Dinucleotide pattern density of *C. cereale* sequence data	90
2.3	Results of the Shimodaira-Hasegawa likelihood ratio test	91
Chapter 3		
3.1	Fungal strains used in this study	120
3.2	Primer sequences used in this study	121
Chapter 4		
4.1	Summary AMOVA table	153
4.2	Pairwise population PhiPT values	154

4.3 Pairwise population N_em values 155

SI4.1 *Colletotrichum cereale* strains used in the multi-locus sequence analysis 156

SI4.2 Isolates of grass-associated *Colletotrichum* species and outgroup taxa used in this study 163

Chapter 5

5.1 Microsatellite loci identified from the enriched genomic DNA of *Colletotrichum cereale* isolate NJ-6340 and PCR primer sequence.. 190

5.2 Isolates of *Colletotrichum cereale* screened using microsatellite primers 192

5.3 Amplification products using the microsatellite primer pairs 193

Chapter 6

6.1 The 55 strains of *Colletotrichum* isolated from grass hosts used in this study ... 227

6.2 *Colletotrichum* isolated from non-graminicolous hosts used in this study 229

6.3 Hyphopodial appresssoria characteristics of the graminicolous *Colletotrichum* 231

Chapter 7

7.1 The 55 strains of *Colletotrichum* isolated from grass hosts used in this study ... 293

7.2 *Colletotrichum* isolated from non-graminicolous hosts used in this study 294

7.3 Species identity of the *Colletotrichum* strains from this study as determined through searches of GenBank using ITS sequences 295

7.4 Representation and characteristics of *Colletotrichum* and *Glomerella* ITS sequences curated by GenBank 296

7.5 Reports of new plant disease caused by species of *Colletotrichum* reported in the journal *Plant Disease* 1997-2007, and the methods used for identification..... 297

List of figures

Figure	Title	Page #
Chapter 1		
1.1	Multi-locus maximum likelihood phylogenetic tree	51
1.2	Split decomposition network	52
1.3	Enlarged splits graph topology	53
1.4	*Colletotrichum cereale* morphological structures	54
1.5	ITS gene tree	55
1.6	HMG gene tree	56
1.7	*Sod*2 gene tree	57
1.8	Herbarium lectotype specimen	58
Chapter 2		
2.1	The *pogo* family DNA transposon *Collect1*	92
2.2	The *Pseudoviridae* family DNA retrotransposon *Ccret1*	93
2.3	The *Metaviridae* family retrotransposon *Ccret2*	95
2.4	The *Colletotrichum cereale* transposon nucleotide base composition (A+T%) and TA/AT ratios	97
2.5	Frequencies (observed/ expected) of dinucleotide pairs in *C. cereale*	99
2.6	*Ccret2*$^{POL2/3}$ phylogenetic tree	101
2.7	*Ccret2*$^{POL2/3}$ network topology	103
2.8	Sliding window analysis	104

Chapter 3

3.1 Map of Pennsylvania, illustrating the origination of the *Colletotrichum cereale* isolates used in this study 122

3.2 Multilocus phylogenetic tree 123

3.3 Transposon Southern blot hybridizations of *Hind*III digested genomic DNA 124

3.4 Schematic tree showing the presence or absence of the transposons evaluated in this study 125

Chapter 4

4.1 North American collection sites sampled in this study 167

4.2 Maximum likelihood phylogenetic tree 169

4.3 *C. cereale* median joining network 170

4.4 W1 gene tree 171

4.5 M72 gene tree 172

4.6 *Sod2* gene tree 173

4.7 ITS gene tree 174

4.8 *Colletotrichum cereale* multilocus phylogenetic tree 175

4.9 Evolutionary hypothesis suggested by this research 176

Chapter 5

5.1 Sequence of plasmid clones containing microsatellite motifs suitable for marker development 194

5.2 Representative alignment of *C. cereale* microsatellite-containing locus with the same locus from the *C. graminicola* genome 202

Chapter 6

6.1 Phylogeny of the falcate-spored graminicolous *Colletotrichum*..... 232

6.2 Cultures of *Colletotrichum cereale*; isolated from cool-season golf course turf ... 233

6.3 Cultures of *Colletotrichum cereale*; isolated from cool-season cereal crops 235

6.4 Cultures of *Colletotrichum cereale*; isolated from various cool-season grasses 236

6.5 Cultures of falcate-spored *Colletotrichum* isolated from warm-season grasses . 238

6.6 Cultures of falcate-spored *Colletotrichum* from non-graminicolous hosts 240

6.7 Photos of *Colletotrichum cereale* hyphopodial appressoria 244

6.8 Photographs of hyphopodial appressoria produced by falcate-spored *Colletotrichum* species from warm-season grasses 248

6.9 Photographs of hyphopodial appressoria produced by falcate-spored *Colletotrichum* species from non-graminicolous hosts 254

6.10 Evolutionary relationships of graminicolous *Colletotrichum* and hyphopodial appressorial characters 260

6.11 Herbarium specimens 261

Chapter 7

7.1 History of *Colletotrichum* systematics 302

7.2 ITS gene tree 303

7.3 Conidia photographs..... 304

Preface: Introduction to the dissertation: Evolution of *Colletotrichum* species inhabiting grasses in diverse ecosystems

P.1 Introduction

One of the most far-reaching intellectual upheavals for the fungal kingdom began in the early 1990s, with the advent of molecular tools that could objectively distinguish minute differences between organisms long considered to lack any discernable variability. The changes wrought by this single technological advance would have been inconceivable even ten years previous. Molecular biology, simply put, set into motion the overthrow of some of the most basic beliefs in the field of mycology, particularly our understanding of fungal evolutionary and population biology. Before the advances of this molecular revolution, fungal evolution and taxonomy were largely descriptive arts, dictated solely by morphological phenotype. But even with small amounts of molecular data, suddenly it became possible to draw connections between unique morphological states produced by seemingly divergent fungi, linking asexual anamorphic "species" with their sexual teleomorphic state. Where once there were individual species described through morphological novelties, now there was an eruption of morphologically cryptic species, uniquely distinguished through their molecular signatures.

Mycology was irrevocably transformed by advancements in DNA sequencing technology, automated data collection and computational advances. Relatively small haploid genomes and advanced sequencing technologies (ABI SOLiD; 454 sequencing) now make possible comparative analyses between genomes of closely related fungal

species, or even between multiple individuals from a single species. But the field is still in the process of moving beyond the narrow confines of typological thinking – an ideology espoused by no living evolutionary biologist – but one that nevertheless continues to permeate the entire fungal kingdom. Even today, 40 years after Whitaker's landmark establishment of the fungal kingdom as an entity distinct from the plant kingdom, fungal taxonomy and nomenclature is still dictated by the tenants of the botanical community. While the evolution of many model systems and economically important fungal groups have been restructured in accordance with phylogenetically-derived species concepts, countless others remained mired in outdated, ill-fitting morphological constraints. The genus *Colletotrichum* is just such a group.

P.2 What drives the evolution of *Colletotrichum* species across diverse grass ecosystems?

Fungi in the genus *Colletotrichum* are destructive pathogens, afflicting over 900 plant species worldwide, including numerous grasses and cereal crops Despite their economic importance and global distribution, the evolutionary processes that shape the genus *Colletotrichum* are poorly understood, and the taxonomic and classification standards have long been considered inadequate and/or inaccurate (Sutton, 1992; Cannon *et al.* 2002). In this dissertation I reconstruct the evolution of grass-inhabiting *Colletotrichum* and relate it to pathogenicity. I have undertaken this research to (1) understand how graminicolous *Colletotrichum* were shaped by evolutionary processes such as population divergence, gene flow, mutation, recombination and speciation; and (2) evaluate whether the lifestyles of these fungi can be correlated with genotypic or genomic signatures, life history or

ecological adaptations. These broad objectives were centered on the application-based model of *C. cereale*, a recently emerged pathogen of the golf course turfgrasses. These studies will span multiple taxonomic levels -- ranging from the supra-specific relationships that form the basis of the graminicolous *Colletotrichum* group, to the genesis of individual species, followed by the differentiation of populations and individuals, and culminating in the diversity of intragenomic populations. For this research, I have chosen to focus on the diversity of two important components of the fungal genome that may themselves have influenced the evolutionary trajectory of the graminicolous *Colletotrichum*: (1) the genes that comprise the fungal mating-type locus and (2) a sample of mobile transposable genetic elements. As part of this work, I compare the utility of various taxonomically informative characters (multi-locus nucleotide sequence data, ITS sequence data, RFLP fingerprints, host range and morphology) within the context of the falcate-spored, graminicolous *Colletotrichum* group to evaluate whether these measures of evolutionary relatedness and taxonomy might prove useful for studies of other *Colletotrichum* groups.

P.4 The genus *Colletotrichum*: Pathogens causing disease worldwide

Filamentous fungi in the genus *Colletotrichum* are found worldwide in association with more than 900 plants, inhabiting countless natural and man-made ecosystems (Cannon *et al*, 2000). The extensive economic losses experienced as a result of *Colletotrichum*-induced plant disease makes this genus one the most successful and important fungal groups. Infection by species of *Colletotrichum* produces anthracnose, a type of plant disease

symptom characterized by the presence of sunken, necrotic lesions on the surface of infected plant tissue.

In addition to their pathogenic lifestyles, many *Colletotrichum* species colonize plant hosts asymptomatically either as mutualists or commensals (Rodriguez and Redman 2008, for review). Non-pathogenic symbiotic *Colletotrichum* colonize living host tissues without inciting any symptoms of infection, and disease induction is either delayed or never occurs. Several *Colletotrichum* species, including *C. dematium*, C. *magna*, *C. musae*, *C. orbiculare*, *C. acutatum*, *C. gloeosporiodes* and *C. cereale* have been isolated from asymptomatic hosts (Sinclair, 1991; Redman *et al*, 2001).

P.3 Taxonomy, classification and species concepts in the genus *Colletotrichum*

Colletotrichum (Corda) is the asexual state assumed by ascomycete fungi of the sexual genus *Glomerella* (Spauld. & H. Schrenk; Lineage: Eukaryota; Fungi/Metazoa group; kingdom Fungi; subkingdom Dikarya; phyla Ascomycota [{Berk} Caval.-Sm]; subphylum Pezizomycotina [O.E. Erikss. & Winka]; class Sordariomycetes [O.E. Erikss. & Winka]; subclass Sordariomycetes incertae sedis; order Phyllachorales [M.E. Barr]; family Phyllachoraceae [Theiss. & Syd]). From phylogenetic analyses of multilocus sequence data, the closest known sister taxa to *Colletotrichum* are mitsporic species of the genus *Verticillium* (Spatafora *et al*, 2006). Traditional *Colletotrichum* systematics relies heavily upon host plant association (Cannon *et al*., 2000); this criterion has been repeatedly shown to be unreliable for the genus; in some instances several species are capable of infecting a single host plant or a single species may be associated with an extremely broad host range

(Freeman, 2000 for review). Morphological characters in the genus may also be systematically uninformative, and are often inadequate to resolve even interspecific relationships with any level of confidence (Sutton 1992). Likewise, molecular investigations of the group conducted to date fail to robustly infer evolutionary relationships between the *Colletotrichum* taxa, with unresolved, minimally supported topologies (e.g., Moriwaki et al., 2002) from which no significant conclusions can be drawn.

References

Cannon, P. F., Bridge, P. D., and Monte, E. 2000. Linking the past, present, and future of *Colletotrichum* systematics. Pp. 1-20 in: *Colletotrichum*: Host specificity, pathology, and host-pathogen interaction. D. Prusky, S. Freeman and M. B. Dickman (eds.). APS Press, St. Paul, MN.

Freeman, S. 2000. Genetic diversity and host specificity of *Colletotrichum* species on various fruits. Pp. 131-144 in: *Colletotrichum*: Host specificity, pathology, and host-pathogen interaction. D. Prusky, S. Freeman and M. B. Dickman (eds.). APS Press, St. Paul, MN.

Johnston, P. R., and Jones, D. 1997. Relationships among *Colletotrichum* isolates from fruit-rots assessed using rDNA sequences. Mycologia 89:420-430.

Moriwaki, J., Tsukiboshi, T., and Sato, T. 2002. Grouping of *Colletotrichum* species in Japan based on rDNA sequences. J. Gen. Plant Pathol. 68:307-320.

Redman, RS, Dunigan, DD and Rodriguez, RJ. (2001) Fungal symbiosis: from mutalism to parasitism, who controls the outcome, host or invader? New Phytol. 151:705-716.

Rodriguez, RJ and Redman, RS. (2008) More than 400 million years of evolution and some plants can't make it on their own: plant stress tolerance via fungal symbiosis. J. Exper. Bot. Doi:10.1093/jxb/erm342.

Sinclair, J.B. (1991) Latent infection of soybean plants and seed by fungi. Plant Dis. 75: 220-224.

Spatafora JW, Sung GH, Johnson D, Hesse C, O'Rourke B, Serdani M, Spotts R, Lutzoni F, Hofstetter V, Miadlikowska J, Reeb V, Gueidan C, Fraker E, Lumbsch T, Lücking R, Schmitt I, Hosaka K, Aptroot A, Roux C, Miller AN, Geiser DM, Hafellner J, Hestmark G, Arnold AE, Büdel B, Rauhut A, Hewitt D, Untereiner WA, Cole MS, Scheidegger C, Schultz M, Sipman H, Schoch CL (2006) A five-gene phylogeny of Pezizomycotina. Mycologia 98(6):1018-28.

Sutton, B. C. 1992. The genus *Glomerella* and its anamorph *Colletotrichum*. Pp. 1-26 *In Colletotrichum*: Biology, pathology and control. J. A. Bailey and M. J. Jeger (eds.). CAB International, Wallingford, U.K.

Chapter 1: Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn

ABSTRACT

Colletotrichum species cause anthracnose diseases on a number of grass hosts and are common inhabitants of many others. They are divided into four species: *C. sublineolum* is pathogenic to *Sorghum* spp.; *C. caudatum* is found on C4 grasses such as indiangrass and big bluestem; *C. falcatum* causes red rot of sugarcane; and *C. graminicola sensu lato* G.W. Wilson is a broadly defined species including isolates that attack maize, wheat, oats, and many forage, turf and amenity grasses of the subfamily Pooideae. In this paper, a combination of hierarchal and non-hierarchal-based analyses was employed to examine evolutionary relationships among the grass-infecting *Colletotrichum* species, with special emphasis on isolates from turf and other grasses in the Pooideae. Reconstructions performed with datasets from over 100 *Colletotrichum* isolates at three variable loci using phylogenetic and network-based methodologies unambiguously supported the taxonomic separation of maize-infecting isolates of *C. graminicola* from the pooid-infecting strains of *Colletotrichum*. To reflect the evolutionary relationships that exist between these distinct lineages, we propose the resurrection of the species name *C. cereale* to describe the pooid-infecting isolates. There was also support for further subdivision of *C. cereale*, but the current data are insufficient to confidently subdivide the species, as there was some evidence of recombination between lineages of this species.

1.1 Introduction

When confronted with fungi that rapidly emerge as destructive pathogens in cultivated plant communities, an understanding of how genetic variation is organized often allows us to reconstruct the sequence of events influencing both the onset and trajectory of disease epidemics. By considering fungal phytopathogen life histories and diversity, it has been possible to deduce the events that have shaped many of the most destructive plant disease outbreaks during the past century. For example, the manifestation of the Dutch elm disease pandemics caused by *Ophiostoma* spp. is known to be characterized by a series of rapid intercontinental migratory events, introgressive gene transfers resulting in novel hybrid genotypes, and, significantly, by the replacement of one pathogen (*O. ulmi*) by a new, more destructive species (*O. novo-ulmi*) (Brasier, 2000 for review). Similarly, by reconstructing phylogenetic relationships among host-specific populations of the rice blast pathogen (*Magnaporthe oryzae*), a recent study suggests that host shifts of the pathogen from non-rice plants to rice crops occurred during the era of rice domestication (Couch et al., 2005). Pursuit of this knowledge is not merely academic: with enough information, plant pathologists have a greater probability of implementing successful disease control measures.

In cereals and grasses, anthracnose disease is caused by four distinct *Colletotrichum* species, each delimited partly by host specificity, but primarily differentiated by morphological features. *Colletotrichum sublineolum* Henn. is pathogenic to *Sorghum* spp.; *C. caudatum* Peck is found on a number of C4 grasses such as indiangrass and big bluestem (*Sorghastrum nutans* (L.) Nash and *Andropogon gerardii* Vitman, respectively); and *C. falcatum* Went causes red rot of sugarcane (*Saccharum officinarum* L.). *Colletotrichum graminicola* (Ces.)

G.W. Wilson is generally considered to inhabit a wide range of hosts including corn (*Zea mays* L.), wheat (*Triticum aestivum* L.), oats (*Avena sativa* L.), and many forage, turf and amenity grasses of the Poaceae subfamily Pooideae (Wilson, 1914); although Sutton formally proposed in 1980 that the species circumscription is only properly applied to *Colletotrichum* on corn (Sutton, 1980). The valid publication of *C. graminicola sensu stricto* Sutton in 1980 rendered illegitimate the application of the name *C. graminicola* to any *Colletotrichum* from hosts other than corn; however, since no diagnoses of a new species taxa for the fungi distributed on pooid grasses has been established, *C. graminicola sensu lato* G.W.Wilson is still routinely, but inappropriately, employed to describe these strains (Baxter et al., 1983).

Historically, *C. graminicola sensu lato* G.W. Wilson was known as a plant pathogen of minor importance (Couch, 1979; Dale, 1963), and in almost all grasses this is still true. But toward the end of the 20th century, two devastating anthracnose disease epidemics caused by *C. graminicola* occurred quite suddenly in North America: first, in corn crops and later, in turfgrasses of the subfamily Pooideae cultivated as golf course greens [especially annual bluegrass (*Poa annua* L.) and creeping bentgrass (*Agrostis stolonifera* L.)]. The recent disease upsurge caused by *C. graminicola* has devastated turfgrasses maintained as golf course greens since the early 1990s and shows no signs of abating. In turf, changing cultivation practices are likely a contributing factor to disease development, as stands of grass are exposed to increasingly harsh management regimes designed to enhance playability for golfers. Factors such as minimal nitrogen fertilization, decreased cutting heights, variability in fungicide efficacy and the increased usage of plant growth regulating chemicals have all been implicated in the enhancement of anthracnose disease levels (Dernoedon, 2000; J. Inguagiato and B. Clarke, *unpublished data*).

As anthracnose disease in turfgrasses attributed to *C. graminicola* became increasingly problematic throughout the 1990s, a number of research groups conducted analyses of molecular genetic diversity or inferred gene genealogies intended to characterize variability in the *C. graminicola* species group. Several conflicting hypotheses were proposed concerning the evolution and diversity of *C. graminicola*, but no theory has generally been accepted as an accurate representation of the species history. Using molecular data, it has been suggested that there is a close connection between *C. sublineolum* and *Colletotrichum* isolates causing anthracnose in annual bluegrass turf (Backman et al., 2000; Horvath and Vargas, 2004); another study supports an association between isolates from creeping bentgrass turf and those from corn (Backman et al., 2000). Alternatively, data from other research groups supports a separate lineage of turfgrass pathogens, diverged from corn and sorghum isolates, which share a single ancestor (Browning et al., 1999; Crouch et al., 2005; Du et al., 005; Hsiang and Goodwin, 2001).

In this study we jointly consider three complimentary datasets to investigate the patterns of evolutionary relationships among *Colletotrichum* isolates responsible for anthracnose disease in turfgrasses, corn, and other grass hosts. We tested the following expectations: (1) *C. graminicola sensu lato* G.W. Wilson is actually an assemblage of divergent lineages, rather than a single cohesive species; and (2) although this fungus appears to reproduce clonally through asexual conidia (Chen et al., 2002), with the sexual state (*Glomerella*) not yet observed in a natural setting, lineages of these fungi may still be structured to some degree by reticulate evolution. To consider these concepts, data from over 100 *Colletotrichum* isolates from grasses and cereals were collected from more than 88 populations throughout the U.S. and Canada. A combination of hierarchal- and nonhierarchal-based approaches was used to rigorously test the limits of species

boundaries and provided compelling evidence of sympatric, species-level phylogenetic divergence between *C. graminicola* pathogenic to maize and *Colletotrichum* populations causing disease in grasses of the subfamily Pooideae, including cultivated C3 turfgrass species. In particular, these analyses demonstrated that *Colletotrichum* isolated from maize and pooid grasses are each more closely related to *C. sublineolum* from sorghum than these two lineages are to one another. To reflect the evolutionary relationships that exist between these distinct organisms, we formally propose to resurrect and revise the species *C. cereale* Manns (Selby and Manns, 1909) to describe the unique lineages isolated from pooid grasses, while *C. graminicola sensu stricto* Sutton uniquely describes *Colletotrichum* from corn hosts. To aid communication throughout this narrative and to emphasize the fundamental distinction to be made between these highly divergent lineages, we will henceforth refer to any *Colletotrichum* specimen isolated from pooid grasses as a member of the *C. cereale* species group, and *C. graminicola* as the species of *Colletotrichum* pathogenic to corn.

1.2 Materials and Methods

1.2.1 Taxon sampling for molecular analysis

A total of 107 specimens of *C. graminicola* isolated from North American grasses (Table 1) were assessed for this study, with the majority of samples (76%) isolated from diseased stands of turfgrass maintained as golf course greens. This sample collection includes specimens from 88 localities, and closely reflects the geographic regions where

anthracnose disease on golf course greens is most prevalent in North America; not all known localities were sampled.

Collected fungi were established in pure culture on potato dextrose agar (Fisher Scientific, Hampton, NH), then single-spore purified to ensure isogenic lines. Cultures were preserved as dehydrated mycelia on Whatman glass fiber at -20 °C. Identification of fungal colonies as *Colletotrichum* was performed using spore and setae morphological characters and later confirmed by performing a BLAST search (Altschul et al., 1990) of the ITS rDNA sequences against the NCBI database.

1.2.2 Outgroup taxa

Phylogenetic relationships within the genus *Colletotrichum* are currently ill-defined, making the choice of an appropriate outgroup to root our phylogenetic tree uncertain based upon previously published hypotheses. Traditional *Colletotrichum* systematics relies heavily upon host plant association (Cannon et al., 2000); this criterion has been repeatedly shown to be unreliable for the genus; in some instances several species are capable of infecting a single host plant (e.g., Johnston and Jones, 1997; Liyanage et al., 1992) or a single species may be associated with an extremely broad host range (Freeman, 2000 for review). Morphological characters in the genus may also be systematically uninformative, and are often inadequate to resolve even interspecific relationships with any level of confidence (Sutton 1992). Likewise, molecular investigations of the group conducted to date fail to robustly infer evolutionary relationships between the *Colletotrichum* taxa, with unresolved, minimally supported topologies (e.g., Moriwaki et al., 2002) from which no significant conclusions can be drawn. Because of the ambiguous state of *Colletotrichum* taxonomy, we constructed a phylogenetic tree using the nucleotide sequence from the ribosomal DNA

(rDNA) internal transcribed spacer region (ITS) nucleotide sequence to empirically select an appropriate species to use as outgroup taxa for this study. BLAST searches of the NCBI database Genbank using the ITS sequence from strains MO-1001178, PA-50005 and NJ-6340 identified a number of *Colletotrichum* species likely to be closely related to *C. graminicola*. Twenty-four sequences were analyzed (AY536046, AF272786, AF489567, AJ301951, AJ301922, AF272783, AF272782, AB042305, AB042304, AY376530, AB196301, AJ301954, AF411770, AJ536231, AF534469, AJ301975, AJ301968, AND AJ311391) along with a representative selection of *C. graminicola* ingroup taxa. The Phyllachoracean fungus *Verticillium albo-atrum* was included as the outgroup. Based upon the resultant 75% consensus tree topology (not shown), four isolates of *C. sublineolum* from sorghum were included in this study, since this species appeared to be a closely related sister taxon to *C. graminicola*. In addition, two isolates of *C. acutatum* J.H. Simmonds were included as outgroup taxa because they are clearly distinct from *C. graminicola* both morphologically and on the molecular level, yet still show a close enough relationship to allow for unambiguous nucleotide sequence alignments.

1.2.3 DNA isolation, amplification and sequencing

Total genomic DNA was extracted from fungal cultures using a standard phenol:chloroform protocol as described (Crouch et al., 2005). Polymerase chain reaction (PCR) products for the internal transcribed spacer (ITS) region of the *ITS1*, *5.8S* and *ITS2* rDNA and the conserved HMG-box of the *MAT1-2* mating type locus (HMG) were generated using published primer pairs and amplification conditions: ITS4 and ITS5 (White et al., 1990) for the ITS region, NcHMG 1 and 2 (Arie et al., 1997) or CgHMG 1 and 2 (Vaillancourt et al., 2000) for the HMG sequence. New primer pairs

were designed to amplify a 625- or 505-bp portion of a single copy manganese-type superoxide dismutase gene (*SOD-2*) (Fang et al., 2002): SOD625F/R (5'GCCCACAGTACATATTGCCTAAGC3' and 5'TCATCCCGGGAGCCAGAAAACCT3') or SOD507F/R (5'ATGGCAGCCTTTCCGTTGAGATAC3' and 5'AGTTGACATGAAGCCACCTACAGC3'). PCR primers were synthesized by Sigma Genosys (The Woodlands, TX), and all amplifications were performed in a Biometra UNO thermoblock (Whatman Biometra, Goettingen, Germany) in 25 μl reactions using 37.5 ng of genomic DNA, 1 U Taq DNA polymerase (Promega, Madison, WI) in 10x PCR buffer, 1.5 mM of $MgCl_2$, 0.2 mM of each dNTP, and 12.5 ng of each primer. *SOD-2* products were amplified from genomic DNA using an initial denaturation at 95°C for 5 min, followed by 40 cycles of denaturating at 95°C for 1 min, annealing at 53°C for 1 min, and extension at 72°C for 1 min; with a final 72°C extension for 10 min. Negative controls were included in all amplifications to check for possible contamination. Amplified fragments were visualized on 0.8% agarose gels, then excised and purified using the Gene Clean III kit (Qbiogene, Irvine, CA). Nucleotide sequences were generated from the forward and reverse strands directly from the amplicon using the corresponding PCR primers and BigDye Terminator cycle-sequencing chemistry (Applied Biosystems, Inc., Foster City, CA) on an ABI 3100 capillary sequencer following the manufacturers protocol but using a quarter of the suggested reaction volume.

Nucleotide sequences were assembled and edited using the Lasergene Sequence Analysis Software package (DNASTAR, Inc., Madison, WI). 318 DNA sequences were deposited in GenBank under the accession numbers DQ126157 to DQ126262 (ITS), DQ131924 to DQ132027 (HMG), and DQ132028 to DQ132051 and DQ133257 to

DQ133340 (SOD2).; the remaining 15 sequences were not reported as they overlapped data already placed by Du et al. (Du et al., 2005) (DQ003109-12, DQ003114, DQ003116-17, DQ002855-59, DQ002826, DQ002862, DQ002865, and DQ002868). Sequences were aligned using the Clustal W algorithm (Thompson et al., 1994), then manually adjusted in Microsoft Word (Microsoft Corp., Redmond, WA). The alignment of protein coding regions was refined according to amino acid sequences and intron positions. Gaps were removed from the dataset, coded as single multi-state characters, then reintroduced as coded data for the phylogenetic analyses.

1.2.4 Phylogenetic analyses

We used both Bayesian maximum likelihood and maximum parsimony methods to infer phylogenetic relationships from each gene region and for the combined data set. Models of evolution and maximum likelihood (ML) parameters were estimated individually for each data set using the Akaike Information Criterion (AIC) (Akaike 1974) as implemented in Modeltest v.3.06 (Posada and Crandall, 1998). Bayesian analyses were performed using MrBayes v.3.0b4 (Huelsenbeck and Rannala, 2004), which allows the different data partitions to be modeled separately using the evolutionary models and informative priors that best describe the data. The optimal model selected under the AIC implemented in Modeltest was specified as prior for each gene. One cold and three incrementally heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC) were started from random trees and run simultaneously for 14,000,000 generations each to sample the phylogenies according to their posterior probabilities. Three replicate runs were performed, and tree topologies and stationarity levels were compared for convergence (Huelsenbeck and Rannala, 2004) to determine the point at which stationarity was achieved in each run, log

likelihood scores were plotted against generation time; the initial generations sampled before convergence were discarded as burn-in. We sampled trees from every 500 generations across all four independent analyses to calculate posterior probabilities for each branch in the maximum likelihood tree. Trees sampled from the posterior distribution were imported into PAUP* v.4.0b10 (Swofford 2000) and, after excluding the burn-in, a 75% majority-rule consensus tree was constructed, with the percentage of samples recovering a particular clade representing that clade's posterior probability (Huelsenbeck and Ronquist, 2003). Bayesian posterior probabilities correspond to the statistical probability that a clade is present in the true tree, given the specified priors, the likelihood model, and the data examined (Huelsenbeck and Ronquist, 2003; Larget and Simon, 1999); a final 75% consensus tree was then generated from all trees in the posterior distribution sampled from the three runs (83,539 trees).

Maximum parsimony (MP) analyses were performed using PAUP* under the heuristic search option, with starting trees obtained from 100 random-addition replicates, and using a tree-bisection reconnection (TBR) branch swapping strategy. Multiple MP trees for any dataset were combined into single strict consensus trees. Nonparametric bootstrap resampling was conducted to evaluate relative levels of support for individual nodes (Felsenstein 1985) using 1,000 bootstrap pseudoreplicates with 100 random additions and TBR branch swapping. MP analyses were conducted for both the individual data sets and the total combined data set.

1.2.5 Estimating recombination and its impact on phylogenetic inference

Although *C. graminicola* is thought to reproduce in nature almost exclusively by clonally generated conidia, populations of this fungus still may be structured to some extent by recombination, either through historical events or through the exchange of genetic material by means of vegetative anastomosis between different individuals. Since the presence of recombination has the potential to produce conflicting phylogenetic signal, and may lead to the recovery of incorrect evolutionary hypotheses, we investigated whether relationships between the *C. graminicola* taxa are more accurately represented by a reticulating network rather than a bifurcating phylogenetic tree topology. First, topologies of individual gene genealogies and the multi-locus gene tree were visually inspected for the presence of incongruence suggestive of recombination. To detect recombination events not reflected in the tree topology, a combination of methods that use substitution patterns and site incompatibility to infer the presence of conflicting signal were employed. After removing indels from the aligned multilocus dataset, sequences were collapsed into unique haplotypes using SNAP Map and SITES version 1.1 (Hey and Wakeley, 1997) launched through SNAP Workbench (Price and Carbone, 2005). Since the majority of the ingroup taxa in our analyses are likely to be recently diverged, all sites violating the infinite sites model, which assumes that the possibility of multiple mutations at a single nucleotide site can be ignored due to extremely low mutation rates, were eliminated. The resultant haplotype dataset was used to identify the presence of reticulating relationships inappropriately represented in an acyclic tree topology; such relationships would be generated by evolutionary processes acting at the population level such as hybridization between lineages or recombination between genes, or homoplasic events such as convergent, parallel or reversing mutations. We used the split decomposition network method, a transformation-based approach which uses distance

parsimony to partition datasets into "splits" of sequences (Bandelt and Dress, 1992). The splits are combined successively, with any incompatible, contradictory groupings introducing a loop into the network to indicate the conflict. We employed the computer program SplitsTree v4.0 b14 (Huson 1998) to visualize reticulation networks in the dataset, with branch support estimated by performing 10,000 bootstrap pseudoreplicates. In the split-graph network, reticulating taxa will be connected by more than one branch, with each connection representing alternative solutions; if the data are perfectly phylogenetic, a bifurcating tree topology will result.

To further explore the boundary between hierarchal and nonhierarchal, tokogenetic relationships in our sample, a site compatability matrix was generated from the haplotype dataset using the SNAP Clade and SNAP Matrix functions of SNAP Workbench. The compatability matrix was used to identify the presence of compatibility/incompatibility among the haplotypes, with any resultant incompatible sites removed from the dataset. The dataset was also evaluated for the signature of recombination events by using Hudson's four gamete test (Hudson and Kaplan, 1985) executed in DnaSP (Rozas et al., 2003); any sites identified as potentially recombinant were also excluded from the matrix. The resultant haplotype dataset, edited of any potentially confounding recombinant or homoplasic sites, was used to perform MP analysis in PAUP* to visualize those portions of the tree topology uniquely characterized by hierarchal, species-level relationships.

1.2.6 Morphological analysis

To determine if the fungi described in the present study as *C. cereale* conform to the morphological description established by Selby and Manns (Selby and Manns, 1909), we

inspected original syntype specimens from the Kansas State Herbarium, presently curated by the New York Botanical Garden Steere Herbarium (NYBG) (NYBG specimen numbers 305598 [type] and 276683 on *Bromus secalinus*; 305599 on *Avena sativa* [type]; 305595 on *Arrhenatherum elatius* [type]; 305596 [type] and 276680 on *Secale cereale*; 305597 [type], 276684 and 276681 on *Triticum vulgare;* 276686 on *Dactylis glomerata*; 276687 on *Phleum pratense;* 276685 on *Agrostis alba*; and unnumbered samples from *Agrostis alba*, *Arrhenatherium elatius*, and *Poa pratensis*) (Fig. 1.9). Morphological examination was also performed on *Colletotrichum* specimens isolated from pooid grasses in our collection (KS-20BIG from *B. inermis*; ONT-00128, NY-16, CT-18, MA-20, MA-21, MA-24, CT-25, and CT-27 from *Agrostis stolonifera*; NJ-6722, NJ-6795, CA-1715, NJ-4990, NH-23, NBR-13, RI-9, CT-8, MA-11, PA-50114, PA-WH3, PA-50623, PA-50231, PA-5005, PA-50111, PA50183, PA-4410, PA-V1, and PA-50002 from *Poa annua*; KS-20DGU and KS-20DGY from *D. glomerata*; NJ-CA1 from *Calamagrostis acutifolia*; and KS-20EVM from *Elymus virginicanus*), with cultures grown on PDA under constant light at 26°C for 7-10 d. Microscopic observations and measurements were made with an Olympus CX40 microscope with bright field or phase contrast illumination. At least 50 conidia from each isolate were measured; dimensions of other fungal structures are given as the range of at least 20 measurements where possible. Photomicrographs were obtained using an Olympus BX41 microscope equipped with a MagnaFire Digital Firewire camera and software (Optronics, Goleta, CA).

1.3 Results

1.3.1 Dataset characteristics and analysis

Because ribosomal DNA genes are maintained at high copy numbers in eukaryotic genomes, non-orthologous copies may potentially be present (e.g., O'Donnell and Cigelnik, 1997), violating the expectation of homology required to accurately reconstruct phylogenetic relationships. In the present study, two lines of evidence support the presence of only orthologous rDNA sequences: nucleotide base calling from the ITS rDNA chromatographs generated by sequencing directly from the PCR products was clean and unambiguous, and all three gene trees were concordant at main branch points and tip clades where phylogenetic signal is adequate to resolve such relationships.

Among members of the *C. graminicola/C. cereale/C. sublineolum* ingroup, few insertion-deletion events were observed in the nucleotide sequence alignments, and none of the gaps introduced rendered the ingroup problematic; ambiguous alignment was restricted to outgroup comparisons. The combined molecular dataset comprised a total of 1,229 nucleotides, with 130 indels coded as standard characters (Table 2). Much of the coded indel data were autapomorphic only with respect to the outgroup taxa and were otherwise parsimony-informative between ingroup taxa; 8 indels of 2-4 bp in length within introns were characteristic of particular monophyletic groupings recovered by the ingroup phylogeny (Figure 6, online supplement).

Given the number of variable characters in the datasets, all three genes appeared to be potentially informative. A preliminary analysis of a 650-bp region of the *TUB-2* gene from the taxa NJ-6340, PA-50005, NJ-6491, MO-1001178 and NY-15182 showed only 1 variable character; therefore it was not included in the phylogenetic reconstruction (*data not shown*). In particular, the 536-bp *SOD-2* sequence consists of 41% parsimony informative characters within the ingroup taxa (ITS=12%; HMG=12%; combined=25%). Individual gene genealogies constructed using strict consensus trees

under parsimony showed visually concordant topologies (Figure 1), although the HMG and ITS sequences were unable to recover some of the groups reconstructed by the more informative *SOD-2* and the combined dataset due to their lower levels of phylogenetic signal. This relative lack of phylogenetic resolution from the *C. graminicola* ITS sequence is consistent with our previously published results (Crouch et al., 2005) and a study conducted by Hsiang and Goodwin (2001); similar reports have been made in other *Colletotrichum* species (Balardin et al., 1999). The lower resolution exhibited by the ITS gene tree may also result to some degree from the extremely large number of equally parsimonious tree topologies recovered during the heuristic searches: 44,034 MP trees were inferred from the ITS sequence data (HMG=4; *SOD-2*=72).

To assess whether the ITS, HMG, and *SOD-2* sequences generate tree topologies that are non-contradictory, we visually compared the individual gene genealogies for evidence of contradictory phylogenetic relationships (Figure 1). Although the incongruence length difference test (ILD; implemented as the partition length homogeneity test in PAUP) is still used in many phylogenetic studies to assess the potential combinability of datasets, several properties of this test are known to generate misleading results (Barker and Lutzoni, 2002; Darlu and Lecointre, 2002; Dolphin et al., 2000; Dowton and Austin, 2002; Sullivan, 1996) particularly when among site rate variation is present (as is the case in the present study), suggesting that the ILD is not a reliable method to strictly determine combinability; we therefore relied on visual inspection to evaluate congruence between the individual gene genealogies. We found only three taxa with inconsistent placement between the gene trees: MA-6722 from *P. annua* (*Sod-2* and combined tree = groups with *C. cereale* clade A isolates; ITS and HMG = groups with *C. cereale* clade B isolates), NJ-8467 from *P. annua* (ITS, HMG, and combined

tree = *C. cereale* clade A; *Sod-2* = *C. cereale* clade B), and NY-15182 from *Z. mays* (HMG, *Sod-2*, and combined = *C. graminicola*; ITS = part of the *C. cereale* lineage). Aside from these three samples, the three gene genealogies all recovered monophyletic groups of *C. sublineolum*, *C. graminicola*, and *C. cereale* taxa, were topologically congruent with one another at these main clades and were also in agreement with population subdivisions recovered through an analysis of three transposable element RFLP patterns (*our unpublished data*); the individual datasets were therefore combined and used for multi-locus analysis.

Analysis of the three gene sequences using the AIC in the computer program Modeltest determined that the evolutionary signature of each sequence was best modeled separately for the Bayesian analyses (Table 3). Although the genes ultimately reconstruct the same relationships between the ingroup taxa, a distinct set of parameters uniquely characterizes the evolutionary processes experienced by each gene. In particular, all three gene sequences showed different rates of variation among different nucleotide sites (among-site rate variation). To correct for the presence among-site rate variation in the dataset, which can cause the number of unobserved, multiple substitutions to be underestimated, the α shape parameter from the gamma distribution was incorporated into the models for each gene (which were established as priors in the Bayesian ML analysis) with α inversely proportional to the amount of among-site rate heterogeneity (when rates are equal, α = infinity). The amount of among-site rate heterogeneity in the *SOD-2* was considerable (α = 0.8737); levels present in the other genes were also estimated to be high (ITS: α= 0.9962; HMG: α = 1.8284).

In the Bayesian likelihood analyses, one run plateaued at 60,000 generations, the second after 119,000 iterations, and the third reached stationarity at 53,000 (negative *ln* likelihood values: run 1: -8785.664 to -8699.69, avg. -8733.853; run 2: -8826.63 to -8740.83, avg. -8740.83; and run 3: -8794.062 to -8658.552, avg. -8697.4453). All three runs converged on the same topology, so all generations sampled from the posterior distribution were combined (83,539 trees) to produce a 75% consensus tree (Figure 2) representing our hypothesis of descent for these taxa. The consensus tree had a *-ln* likelihood score of -8746.140; only clades with relatively strong support (>75% support for nodes from the posterior distribution) were retained.

1.3.2 Phylogenetic relationships

All phylogenetic analyses, separate and combined, using either the complete dataset, the dataset reduced to unique haplotypes, or the haplotype dataset reduced to strictly compatible sites (83 total sites) resolved three distinct, well-supported groups: (1) a monophyletic group composed of the samples isolated from Pooideae grasses (clades A & B, Figure 2), (2) a clade consisting of the sorghum-derived samples (*C. sublineolum*, clade C) and (3) a group composed only of maize-derived taxa (clade D). In the *SOD-2* and combined tree a further subdivision was observed: the *C. cereale* clade subdivided into unique lineages (clades A and B). *C. cereale* clade B further diverged into two smaller groupings. These groups were supported by both bootstrap values and posterior probabilities greater than 95%. The phylogenetic trees recovered under both parsimony and Bayesian likelihood optimality criteria are in agreement at all of the main clades. Figure 2 shows the 75% consensus tree constructed from the 83,539 trees sampled from the Bayesian posterior distribution; both the posterior probabilities and bootstrap support

values are plotted at the nodes (the MP tree is not illustrated separately due to topological congruence with the ML phylogeny; groups not supported by at least a value of 0.75 by both bootstrap and posterior probabilities were collapsed). The results of the molecular phylogenetic analysis reconstructs a species history characterized by the presence of morphologically similar sister species: *C. graminicola*, which is pathogenic to corn, and a group of *C. cereale* lineages found on a wide range of pooid grasses, including cultivated C3 turfgrasses.

A much less well-supported association between *C. graminicola* and *C. sublineolum* was consistently recovered in all of the phylogenetic analyses, with these two species always forming a monophyletic group, separate from the *C. cereale* group. This association between the cereal-derived lineages was not, however, entirely well supported, with a posterior probability of only 0.63 in the Bayesian consensus tree. Analysis of the haplotype dataset using the split decomposition network reconstruction method (Figure 3) also estimates a close relationship between *C. graminicola* and *C. sublineolum*, but confidence in a recent common ancestry for these taxa was extremely low (bootstrap=56).

1.3.3 Reticulation and incompatibility in the dataset

A total of 57 unique multilocus haplotypes were identified from the ingroup sample; with 611 nucleotides distinctively establishing the haplotypes after the removal of introns and any positions violating the infinite sites model. The split-decomposition splits-graph used to visualize the presence of reticulate evolution in the haplotype dataset (Figures 3 and 4) recovered tree-like relationships between the main species clades also estimated by the phylogenetic investigation: *C. sublineolum* from sorghum, *C. graminicola* from corn, and the *C. cereale* species group from the Pooideae subfamily grasses. Within the clades, on the

intraspecific level a reticulating network of relationships clearly emerged: a single loop was present among the *C. graminicola* haplotypes (5 haplotypes, 23 individuals), a single loop connected the *C. sublineolum* haplotypes (3 haplotypes, 4 individuals), and more complex patterns were found individually among the *C. cereale* lineages (Figure 4). The two main *C. cereale* phylogenetic clades A and B were inferred in the splits-graph, with reticulate lineages occurring only below the level of these groups. *C. cereale* clade A consisted of 27 unique haplotypes from 59 individuals, all radiating from haplotype 5 (H5), which, given its central position in the splits-graph, is likely the ancestral haplotype for the *C. cereale* lineages. This interpretation is further supported by the observation that H5 contained the largest number of *C. cereale* samples (22), and was widely distributed throughout the entire range of our sample. The majority of the haplotypes were tightly interrelated in clade A, with only a few divergent lineages. Based upon its position within the splits-graph, clade A haplotype 47 also appeared to be a central lineage among the group, and served as the connecting point between *C. cereale* Clades A and B. Haplotype 46, represented by *C. cereale* sample NJ-8467 from *P. annua*, was positioned directly between clades A and B, which is consistent with the observed conflict between gene genealogies for this isolate (ITS, HMG, and combined=clade A; *SOD-2*=clade B). Collectively, these data suggest that this haplotype is a hybrid between the two *C. cereale* lineages. Haplotype 3, consisting of *C. cereale* isolate MA-6722, while similarly characterized by conflicting gene trees in the phylogenetic analyses (*Sod-2* and combined tree=clade A; ITS and HMG=clade B), appears in the splits-graph to be a highly divergent taxon, but uniquely a member of the clade A lineage.

The haplotypes of *C. cereale* clade B, although numerically less prevalent in our sample, were nevertheless a much more genetically diverse assemblage, with 22

haplotypes resulting from 27 isolates interwoven into a complex pattern of cycles. While the phylogenetic analyses support the split of Clade B into two distinct lineages, the splits-graph was unable to recover the subdivision, suggesting that these taxa are all of the same species.

Analysis of the haplotype dataset by means of the compatibility matrix and Hudson's four gamete test identified the majority of characters as being incompatible; these characters were subsequently removed from the dataset. The resultant 83 character dataset was used to further investigate species-level boundaries through maximum parsimony phylogenetic analysis. This reduced MP phylogeny (not shown) confirmed the splits-graph topology: *C. sublineolum* from sorghum, *C. graminicola* from corn, and *C. cereale* from the Pooideae grasses are each individual species; however, the species level divergence of *C. cereale* Clades A and B found in all other analyses was not recovered.

1.4 Taxonomy

Based upon the unique patterns of fixed nucleotide differences in DNA sequence data at the *ITS1/5.8S/ITS2*, *MAT1-2* and *SOD-2* nuclear loci, it is evident that *Colletotrichum* isolated from the pooid grasses examined in this study form a species taxon distinct from *C. graminicola*, which is limited to the fungus pathogenic to maize host plants. Based upon these nucleotide data, we formally propose to resurrect and emend the species *C. cereale* and to emend the species descriptions of *C. graminicola* and *C. sublineolum*.

3.4.1 ***Colletotrichum cereale*** Manns, Agr. Expt. Sta. Bul. 203:207 (1909).

Teleomorph: Unknown

Emended description: Inhabits grasses of the subfamily Pooideae; pathogenic and sometimes particularly aggressive on *Poa annua*, *P. pratensis*, *P. sapina*, *Agrostis stolonifera*, *A. canina*, *Lolium perenne*. *C. cereale* is uniquely described by the following fixed nucleotide characters (positions are relative to the nucleotide sequence alignment of the epitype specimens, Figure 6): *ITS1/5.8S/ITS2* positions 15 (C), 39 (T), 56 (C), 98 (C), 104 (C), 117 (G), 145 (C), 165 (A), 178 (T), 409 (A), 517 (T), 522 (C); *MAT1-2* positions 2 (A), 54 (C), 79 (A/T), 87 (G), 97 (A), 101 (C), 109 (A), 123 (T), 129 (G), 131 (A), 154 (C), 163 (C), 172 (A), 199 (A); and *SOD-2* positions 35 (C/T), 45 (A), 47 (C), 53 (A), 58 (A), 61 (G), 70 (C), 73 (A), 77 (T), 88 (C), 119 (G), 127 (C), 133 (C), 136 (G), 142 (C), 151 (C), 169 (A), 175 (C), 205 (C), 211 (C), 226 (G), 244 (A), 247 (C), 259 (C), 262 (C), 268 (C), 280 (T), 292 (G), 332 (G), 355 (C), 373 (G/A), 376 (C), 382 (C), 386 (G), 394 (T), 403 (C), 451 (C), 454 (C), 455 (G), 500 (T), 501 (G), 512 (C), 518 (T), 519 (T), 520 (G), 521 (G), 522 (A), 525 (G), 534 (A), 537 (C), 541 (C), 545 (T), 546 (T), 547 (C), 551 (C), 558 (T), and 571 (G).

Colonies on PDA under constant illumination highly variable; usually form dark mat of tight setae masses across agar surface; commonly heavy conidia accumulation gives orange cast to brown/black culture. Some cultures thin layer of hyphae along agar surface, copious amounts of conidia give orange appearance. Some cultures possess fluffy aerial mycelium growing over setae, producing gray appearance; generally as colonies age mycelia overtakes entire culture surface. Hyphae septate, normally hyaline, sometimes dark brown when present at the base of setae, 1.0-6.5 μm, often guttulate. Conidia falcate or fusiform, apices acute, individually hyaline but appear salmon-orange en mass, may be mono-, bi-, or up to seven-guttulate or oil drops may be absent from the cytoplasm, measuring 6.0-33.8 μm x 2.2-6.3 μm with an average of 23.3 μm x 3.4 μm. Germinating conidia form single or infrequently two hyaline germ tubes that terminate in dark

brown/black appressoria, sometimes appressoria absent; germ tubes separated from appressoria by septa (Figure 5), occasionally appressoria forms directly from conidia. Appressoria rounded and smooth or irregular or lobate or multi-lobate, measuring 8.5-11.6 μm x 6.5-10.2 μm. Setae develop from dark-brown, tight masses of hyphae (Figure 5); in culture dense, rounded masses of setae form, (Figure 5) sometimes partly or completely covered in thick deposition of conidia. Individual brown-black setae separated from hyphae by septa. Setae base swollen or not swollen, irregularly septate with up to 7 septa, measuring 32-120 μm x 6-8 μm at base, tapering at tip.

Type specimen: Examination of herbarium syntype specimens confirmed the presence of acervuli on the grass hosts, in association with black-brown setae that are characteristic of *Colletotrichum*. Very few conidia were observed; those that were present were somewhat shriveled and without cytoplasmic contents, falcate in shape, measuring 20-25 μm x 3 μm. Morphological comparison of the syntypes of *C. cereale* and the published descriptions and illustrations against our recent samples obtained from pooid grasses confirmed that the *Colletotrichum* isolated from pooid grasses in the present study are within the range of morphological variation consistent with the *C. cereale* species description. Although we attempted to extract DNA from the herbarium material, we were not successful, as the fungal tissue was in close association with the plant tissue, and relatively little fungal tissue was present. Attempts to revive the specimens on PDA were also unsuccessful, therefore comparisons of contemporary *Colletotrichum* strains with syntype specimens are necessarily confined to morphological estimations. Four of the New York Botanical Garden specimens were catalogued in the herbarium database as type specimens: (although the published description never makes such a specification): 305598 on *Bromus secalinus*, 305599 on *Avena sativa*, 305595 on *Arrhenatherum elatius*, 305596

on *Secale cereale*, and 305597 on *Triticum vulgare*. A holotype was not designated; we therefore establish 305598 (*B. secalinus*) as the lectotype for the species, and 305599, 305595, 305596 and 305597 are specified as paratypes. To facilitate species interpretation using the molecular characters described in this study, we designate KS-20BIG, NJ-6795, PA-5062-3 and NJ-4990 as epitypes; cultures of the epitype strains have been deposited in the American Type Culture Collection, Manassas, VA, U.S.A. (ATCC) and the Centraalbureau voor Schimmelcultures, Utrecht, Netherlands (CVS).

Comments: *C. cereale* also possesses several fixed autapomorphic characters that serve to uniquely distinguish the species from *C. graminicola* (positions are relative to the nucleotide sequence alignment of the epitype specimens, Figure 6): *ITS1/5.8S/ITS2* positions 54 (no base), 69-72 (no bases), 100 (C), 111 (A), 115 (G), 121 (no base), 158 (no base), 404-406 (no bases), 525 (no bases); *MAT1-2* positions 115 (no base), 120 (G); and *SOD-2* positions 18-19 (AA), 25-27 (no bases), 29 (T), 56 (A), 60 (no base), 78 (no base), 82 (no base), 85 (T), 503 (no base), 508 (T), 562 (A), and 567 (no base).

In general, *C. cereale* is morphologically very similar to *C. graminicola*. Two continuous morphological characters (conidia length and hyphopodium area) have recently been suggested to approximately differentiate between *C. graminicola* and *C. cereale* (Browning et al., 1999; Crouch et al., 2005), however, the variability and overlap of range in these types of quantitative characters precludes their ability to uniquely and consistently diagnose the taxa (e.g. Sutton 1980). Conidia in these two species are sometimes different in length, with the conidia of *C. graminicola* are typically, but not always, larger than those from *C. cereale* (Backman et al., 1999; Browning et al., 1999; Du et al., 2005; Selby and Manns, 1909; our data): overall, *C. cereale* conidia grown on PDA measured 6.0-33.8 μm, while *C. graminicola* conidia ranged from 13.0-44.0 μm.

Hyphopodium (mycelial appressoria) area is also notably different in the two species (Browning et al., 1999; Du et al, 2005) with the hyphopodium area of *C. graminicola* on average, significantly larger than those of *C. cereale*, but this continuous character also is present in overlapping ranges (*C. cereale* 63.8-315.6 μm^2 and *C. graminicola* 136.7-1027 μm^2 (12, 25, 61).

1.4.2 ***Colletotrichum graminicola*** (Ces.) G.W. Wilson, Phytopathology 4:110 (1914) (as "*graminicolum*").

- ≡ *Dicladium graminicola* Ces., Flora 35: 398 (1852) (as "*graminicolum*").
- ≡ *Vermicularia graminicola* (Ces.) Westd., Bull. Acad. Roy. Brux. 12: n. 7 (1861).
- ≡ *Steirochaete graminicola* (Ces.) Sacc. Syll. Fung. 4: 316 (1886).
- ≡ *Colletotrichum zeae* Lobik, Trudy severo-kavkazskogo Instituta Zashchity Rastenii 1(2): 39 (1933).
- ≡ *Colletotrichopsis graminicola* (Ces.) Muntaola, Rev. Argent. Agron. 19: 220 (1952).
- ≡ *Colletotrichum graminicola* f. sp. *zeae* Messaien, Lafon & Malot, Ann. Epiphyt., ser. C., 10: 454 (1959).

Teleomorph: ***Glomerella graminicola*** Politis, Mycologia 67: 56-72 (1975).

Emended description: Parasitic on *Zea mays*. *C. graminicola* is uniquely distinguished by the following synapomorphic fixed nucleotide characters (positions are relative to the nucleotide sequence alignment of the epitype specimens, Figure 6): *ITS1/5.8S/ITS2* positions 15 (T), 39 (C), 56 (T), 92 (G), 98 (C), 99 (A), 104 (T), 117 (C), 145 (T), 165 (G), 178 (C), 409 (T), 517 (C), 522 (T); *MAT1-2* positions 2 (T), 54 (T), 79 (G), 87 (A), 97 (G), 101 (T), 109 (G), 123 (C), 129 (A), 131 (C), 154 (T), 163 (T), 172 (G),

199 (G); and *SOD-2* positions 35 (A), 45 (G), 53 (T), 58 (T), 61 (A), 70 (G), 73 (G), 77 (C), 88 (T), 119 (A), 127 (T), 133 (T), 136 (A), 142 (G), 151 (T), 169 (G), 175 (A), 205 (T), 211 (A), 226 (T), 244 (T), 247 (T), 259 (T), 262 (G), 268 (A), 292 (C), 332 (A), 255 (T), 373 (T), 376 (T), 382 (T), 386 (A), 394 (C), 403 (T), 451 (T), 454 (T), 455 (A), 500 (C), 501 (A), 512 (T), 518 (A), 519 (A), 520 (A), 521 (A), 522 (G), 525 (A), 534 (G), 537 (A), 541 (T), 545 (A), 546 (A), 547 (T), 551 (T), 558 (C) and 571 (A).

Epitype establishment: The lectotype established for this species was examined by Sutton (58) from *Z. mays* (IMI83255). To facilitate species interpretation using the molecular characters described in this study, we designate MO-100178 from *Z. mays* (also known as *C. graminicola* sample M1.001) as an epitype; a culture of this specimen has been deposited in the ATCC and the CVS.

Comments: The reader is referred to any of several excellent morphological studies of *C. graminicola sensu stricto* Sutton that have been published (Baxter et al., 1983; Du et al., 2005; Sutton 1965; Sutton 1966; Sutton 1968; Sutton 1980; Sutton 1992). In general, *C. graminicola* is morphologically quite similar to *C. cereale* except for the two continuous quantitative conidial and hyphopodial characters noted above, but the overlapping range of these morphological characters precludes their usage to consistently and uniquely diagnose these species.

C. graminicola also possesses several fixed autapomorphic characters that can serve to uniquely distinguish the species from *C. cereale* (positions are relative to the nucleotide sequence alignment of the epitype specimens, Figure 6): *ITS1/5.8S/ITS2* positions 54 (T), 69-72 (TCCG), 100 (no base), 111 (no base), 115 (no base), 121 (G), 158 (A), 404-406 (GTA), 525 (C); *MAT1-2* positions 115 (C), 120 (no base); and *SOD-2* positions 18-19 (no

bases), 25-27 (AAC), 29 (no base), 47 (no base), 56 (no base), 60 (T), 78 (C), 82 (C), 85 (no base), 500 (C), 508 (no base), 562 (no base), and 567 (C).

1.4.3 ***Colletotrichum sublineolum*** Henn. *apud* Kabat & Bub. Fungi imp. exs. 186 (1905) (as "*sublineola*").

- ≡ *Colletotrichum graminicola* f. sp. *sorghi* Messaien, Lafon & Malot, Ann. Epiphyt., ser C., 10: 456 (1959).
- ≡ *Colletotrichum graminicola* var. *zonatum* Rajasab & Ramal., Curr. Sci. 50(1): 34 (1981).

Teleomorph: Unknown

Emended description: Parasitic on *Sorghum bicolor* and *Sorghum halapense*. *C. sublineolum* is uniquely distinguished by the following synapomorphic fixed nucleotide characters (positions are relative to the nucleotide sequence alignment of the epitype specimens, Figure 6): *ITS1/5.8S/ITS2* positions 39 (T), 67 (T), 93 (C), 98 (T), 99 (C), 105 (C), 112 (C), 131 (G) 132 (A), 133 (G), 135 (A), 136 (T), 137 (A), 139 (G), 142 (A), 143 (A), 147 (T), 178 (T), 417 (T), 419 (A), 420 (C); *MAT1-2* positions 2 (A), 54 (C), 66 (T), 77 (A) 79 (A) 87 (G), 88 (A), 95 (G), 106 (G), 130 (A), 145 (A), 154 (C), 163 (C), 181 (T), 202 (T); and *SOD-2* positions 40 (C), 42 (T), 44 (T), 48 (T), 88 (C), 115 (A), 125 (G), 142 (C), 148 (A), 151 (C), 166 (T), 175 (C), 184 (T), 187 (A), 205 (C), 211 (C), 226 (G), 244 (A), 268 (C), 332 (G), 355 (C), 367 (A), 373 (G), 376 (C), 382 (C), 386 (C), 403 (C), 408 (T), 455 (G), 457 (C), 460 (A), 495 (A), 498 (G), 504 (G), 505 (G), 509 (T), 512 (G), 518 (G), 519 (G), 524 (A), 530 (G), 534 (A), 537 (T), 538 (C), 541 (G) and 563 (A).

Epitype establishment: To facilitate species interpretation using the molecular characters described in this study, we designate S3.001 from *S. bicolor* as an epitype; a culture of this specimen has been deposited in the ATCC and the CVS.

Comments: *C. sublineolum* also possesses several fixed autapomorphic characters that serve to uniquely distinguish the species from *C. graminicola* (positions are relative to the nucleotide sequence alignment of the epitype specimens, Figure 6): *ITS1/5.8S/ITS2* positions 25 (T), 69-72 (no bases), 100 (C), 101-102 (TC), 107 (no bases), 111 (A), 113-115 (GGG), 119-120 (CG), 138 (A), 414-416 (no bases), 514 (T), 522 (no base); *MAT1-2* positions105 (C), 116 (no base), 120 (G); and *SOD-2* positions 35 (no base), 37 (no base), 39 (A), 47 (T), 50 (no base), 52 (no base), 503 (no base), 508 (G), and 544-546 (GGA).

We have not listed a teleomorph state for the species; in a review of the species taxonomy Sutton suggests that the teleomorph may be *G. cingulata* var. *sorghicola* Saccas (Sutton 1992).

It is worth emphasizing that although *C. graminicola* is still frequently employed to describe *Colletotrichum* from sorghum, Sutton's 1980 emendment of the species description rendered such application of the name illegitimate. *C. sublineolum* is the validly established taxon describing *Colletotrichum* associated with sorghum; this distinction has been repeatedly confirmed through morphological examinations (Baxter et al., 1983; Browning et al., 1999; Du et al., 2005; Sutton 1966; Sutton 1980; Sutton 1992), host range and pathogenicity testing (Backman et al., 1999; Browning et al., 1999; Dale 1963; Jamil and Nicholson, 1987; LeBeau 1950) and molecular analyses (Browning et al., 1999; Crouch et al., 2005; Du et al., 2005; Hsiang and Goodwin, 2001; Moriwaki et al., 2002; Randhir and Hanau, 1997; Sherriff et al., 1995; Sreenivasaprasad et al., 1996; Vaillancourt and Hanau, 1991; Vaillancourt and Hanau, 1992; this study).

1.5 Discussion

The purpose of this study was to explore the evolutionary history of the *C. graminicola* lineages that cause anthracnose disease in turfgrasses, in particular, we sought to identify the point at which interspecific boundaries have been erected between taxa, defining the extreme limits of gene flow and population-level processes.

1.5.1 Species boundaries

The phylogenetic analyses performed in this study generated an extremely well supported hypothesis of evolutionary descent for the taxa currently recognized as *C. graminicola sensu lato* G.W. Wilson and conclusively establishes that there are two species within this circumscription. Although we employed several methods to identify reticulate evolution among the taxa, the evidence is overwhelmingly in favor of a long-standing biological isolation between these two distinct, monophyletic groups.

In light of the pattern of fixed nucleotide differences and the unique host plant associations observed between lineages of *Colletotrichum* from corn and from those samples obtained from the C3 grasses in the subfamily Pooideae, the species level phylogenetic divergence present in *C. graminicola sensu lato* G.W. Wilson is unambiguous. We therefore propose to resurrect the species *C. cereale* to describe *Colletotrichum* from pooid grasses and *C. graminicola* (Ces.) G.W. Wils. *sensu stricto* Sutton 1980 is reserved for the fungus pathogenic to corn, with each species uniquely characterized by their molecular identities at the *ITS1/5.8S/ITS2*, *MAT1-2* and *SOD-2* loci and their host range. These physiological host range limitations are well documented in the literature (Ali 1962;

Backman et al., 1999; Browning et al., 1999; Dale 1963; Jamil and Nicholson, 1987; Kemp et al., 1991; LeBeau 1950; LeBeau et al., 1951; Messiaen et al., 1959; Minussi and Kimati 1979; Zeiders 1987; Zwillenberg 1959) and correspond with the emended phylogenetic species described in this paper. Since *C. cereale* likely has a much more extensive host range as a pathogen than that which is explicitly described in this manuscript, we emphasize that it is the pattern of fixed differences on the molecular level and the association with grasses of the Pooideae that serve to uniquely distinguish this species, rather than strict pathogenicity criteria. This is an important distinction: *C. cereale* transcends its ability to induce anthracnose disease symptoms in a given host plant, since the species is capable of inhabiting many pooid grasses without inducing disease in the host.

Even in our most conservative analyses, where only the most reliable, unconflicted nucleotide characters are considered, these species fully meet the criteria of the genealogical concordance phylogenetic species concept (Taylor et al., 2000), which is capable of efficiently and objectively marking species boundaries in asexually reproducing organisms such as *Colletotrichum* that defy characterization based upon morphological or reproductive criteria. The specific designation of these groups is dictated by taxonomic priority: *C. graminicola*, as *Dicladium graminicolum*, was first described by Cesati (Cesati 1852) in 1852 from the stems of corn and barnyard grass (*Echinochloa crus-galli*); the first formal description of *Colletotrichum* from a pooid grass did not occur until 1909 (Selby and Manns, 1909) when Selby and Manns first described *C. cereale*. Most of the taxonomic uncertainty surrounding the circumscription of *C. graminicola* completely bypassed members of the genus found in association with pooid grasses and instead concentrated on those strains responsible for economically important levels of disease in the host plant:

corn, sorghum and sugarcane. Even Sutton, the recognized authority in *Colletotrichum* systematics, has evaluated only two pooid strains as *C. graminicola* (from *Avena sativa*) using conidial measurements; herbaria material inspected from *Bromus* spp., *Calamagrostis epigeoios*, *C. villosa*, *C. neglecta, Lolium perenne*, and *Poa annua* were accepted by Sutton to be *C. dematium* (since the morphological structures examined were not capable of making such a distinction between these taxa rather than by error or intention); and the study of Selby and Manns that served to erect *C. cereale* was not cited or discussed in any of Sutton's publications (Sutton 1965; Sutton 1966; Sutton 1968; Sutton 1980; Sutton 1992).

Because our current research goals are focused on the population dynamics of *Colletotrichum* from pooid grass hosts rather than conducting an expansive taxonomic survey of the genus, we have not sampled *Colletotrichum* from the other C4 grasses evaluated by Wilson (*Panicum* spp., *Echinochloa crus-galli*), leaving the species status of these *Colletotrichum* still undefined; Sutton's 1980 treatment of *C. graminicola* precludes the legitimate use of the name to describe these taxa. In a review of the taxonomy of the genus *Colletotrichum* in 1992 (Sutton 1992) Sutton introduced the idea that a still unnamed species closely related to *C. falcatum* occurs on grass hosts other than sugarcane, basing his suggestion upon earlier morphological studies of samples from several "miscellaneous" C4 grass genera of the Panicoideae including *Andropogon, Digitaria, Echinochloa, Eleusine, Eragrostis, Miscanthus, Panicum,* and *Rottboellia* (Sutton 1965). Molecular data presented in two recent phylogenetic analyses (Browning et al., 1999; Du et al. 2005), confirms Sutton's hypothesis, providing additional evidence that a new species will need to be erected to properly describe the *Colletotrichum* associated *Echinochloa* spp., since these taxa do not phylogenetically conform to any of the species now described from graminaceous host plants.

One particular question that remains unresolved is the relationship between *C. graminicola* and *C. sublineolum*: have they recently diverged from a common ancestor -- an intriguing premise that continually arises in all of the evolutionary inferences, albeit with minimal support. For several years these two species, along with *C. falcatum* and *C. caudata*, were considered to be varietal forms of a single species (Arx 1957; Messiaen et al., 1959; Minussi and Kimati, 1979), and the unique status of *C. sublineolum* is still rejected by many researchers of sorghum anthracnose (e.g. Valero et al., 2005). Is this association between species real, or is it an artifact of ancestral polymorphisms? To fully consider this possibility, a substantially larger portion of the genome should be sampled, with an extended sampling from these taxa along with their sister species from other C4 grass hosts.

1.5.2 Unresolved, potentially species-level divergences

We envision that *C. cereale sensu lato* Crouch, Clarke, and Hillman may eventually need to be further subdivided -- minimally into two species, and potentially into as many as four distinct species based solely upon this limited sampling. We regard the description of *C. cereale* to be a species group rather than a single homogeneous species, but the data from the present study are inadequate for the purpose of formally proposing subdivision of the group. Both ML and MP phylogenetic inferences and the splits-graph strongly support the divergence of *C. cereale* clades A and B, but the phylogeny of the dataset reduced of all potentially incompatible sites does not recover these lineages as independent entities. While this discrepancy is likely due to the elimination of a great many of the informative characters from the analysis rather than continued gene flow, we cannot at this time confidently suggest the establishment of two unique species based upon these data alone,

since the eliminated characters may have yielded erroneous species phylogenies. It is not surprising that a great deal of potentially misleading data can be found when considering a group of organisms so close to the population level, as any number of processes can result in inconsistent, misleading conclusions due to incomplete lineage sorting, recombination, and hybridization. We expect that the application of several analytical techniques -- vegetative compatibility, host range analysis, pathogenicity and virulence -- will be required before these unique groups can be defined in a substantive, biologically meaningful manner. This work merely represents the first step in an ongoing process; nevertheless, it does provide a vital framework for future experimental work.

1.5.3 The evolutionary history of *C. cereale*.

In any phylogenetic analysis, the assumption that only a single phylogeny underlies the evolution of the population sampled is violated by the presence of recombination. Recombination is a truly creative force in organismal biology, but has the unfortunate side effect of completely confounding the accurate estimation of phylogenies when present (Carbone and Kohn, 2004; Posada and Crandall, 2001; Posada and Crandall, 2002 for reviews). This is particularly true in sexual species and, as our data will attest, in putatively asexual entities like those found in the genus *Colletotrichum*. It is evident that although *C. cereale* may for the most part propagate in an asexual fashion, recombination between taxa has occurred, as independently estimated by the split-decomposition network and compatibility matrix analysis; results from transposon RFLP fingerprinting assays also confirm the existence of mosaic genotypes in the species group (J. Crouch, B. B. Clarke, and B. I. Hillman, *unpublished data*). Although the three individual gene genealogies considered in this study were topologically congruent,

suggesting a clonal, non-recombining species history, overall our data provide evidence for the presence of recombination that is not reflected in the tree topologies. Many experts in the field of evolutionary biology recommend using a combination of assays to detect recombination in order to maximize the possibility of identifying recombination with a minimum of false positives (Posada et al., 2002 for review). In our research, three separate analyses detected the potential for recombination between *C. cereale* Clades A and B; we believe these data are sufficient to recommend further analysis before establishing these clades as phylogenetic species when they may in fact only represent divergent populations.

Regardless of whether one accepts the clades of *C. cereale* as individual species or simply highly divergent populations, there are undeniably two unique lineages of this fungus causing disease in turfgrasses maintained as golf course greens. The clade A isolates are geographically widespread, and were isolated from numerous turfgrass and non-cultivated C3 pooid grass species; they are the numerically dominant form of *C. cereale* in the sample discussed in this manuscript and in our entire culture collection. To date, over 200 additional specimens of *C. cereale* in our culture collection have been identified as part of clade A through the application of a PCR-based screening protocol using lineage-specific transposable element primers (J. Crouch, B. B. Clarke, and B. I. Hillman, *unpublished data*). In contrast, the clade B lineage is quite rarely encountered, and appears to have a much more restricted host range and geographic distribution. Additionally, there are several hints of an association between host plant derivation and the *C. cereale* clade B lineage: the phylogenetic tree topology in particular shows an almost perfect division between *A. stolonifera* and *P. annua* isolates. Whether the peculiarities of

clade B with respect to distribution and host range are based on real phenomena or have arisen due to sampling bias will require further study.

One of the most important conclusions generated by these analyses is the identification of the very common and widely distributed H5 ancestral haplotype in the *C. cereale* clade A population. The proposition that all North American *C. cereale* lineages are ultimately derived from the H5 haplotype, regardless of their host-plant derivation, could have very important implications for the development of effective disease control strategies.

1.6 References

Aikaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control 19:716-723.

Ali, M. M. 1962. Factors influencing pathogenicity of three isolates of *Colletotrichum graminicola* on wheat. Mycopathol. et Mycol. App. XIX:161-166.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Arie, T., Christiansen, S. K., Yoder, O. C., and Turgeon, B. G. 1997. Efficient cloning of ascomycete mating genes by PCR amplification of the conserved *MAT* HMG box. Fungal Genet. Biol. 21:118-130.

Arx, J. A. v. 1957. Die arten der gattung *Colletotrichum* Cda. Phytopath. Z. 29:413-468.

Backman, P. A., Landershoot, P. J., and Huff, D. R. 1999. Variation in pathogenicity, morphology, and RAPD marker profiles in *Colletotrichum graminicola* from turfgrasses. Crop Sci. 39:1129-1135.

Balardin, R. S., Smith, J. J., and Kelly, J. D. 1999. Ribosomal DNA polymorphism in *Colletotrichum lindemuthianum*. Mycol. Res. 103:841-848.

Bandelt, H. J., and Dress, A. W. 1992. Split decomposition: A new and useful approach to phylogenetic analyses of distance data. Mol. Phylogenet. Evol. 1:242-252.

Barker, F. K., and Lutzoni, F. M. 2002. The utility of the Incongruence Length Difference Test. Syst. Biol. 51:625-637.

Baxter, A. P., van der Westhuizen, G. C. A., and Eicker, A. 1983. Morphology and taxonomy of South African isolates of *Colletotrichum*. South African J. Bot. 2:259-270.

Brasier, C. M. 2001. Rapid evolution of introduced plant pathogens via interspecific hybridization. BioScience 51:123-133.

Browning, M., Rowley, L. V., Zeng, P., Chandlee, J. M., and Jackson, N. 1999. Morphological, pathogenic,and genetic comparisons of *Colletotrichum graminicola* isolates from *Poaceae*. Plant Dis. 83:286-292.

Cannon, P. F., Bridge, P. D., and Monte, E. 2000. Linking the past, present, and future of *Colletotrichum* systematics. Pp. 1-20 in: Colletotrichum: Host specificity, pathology, and host-pathogen interaction. D. Prusky, S. Freeman and M. B. Dickman (eds.). APS Press, St. Paul, MN.

Carbone, I., and Kohn, L. M. 2004. Inferring process from pattern in fungal population genetics. Pp. 1-30 in: Fungal Genomics, Applied Mycology and Biotechnology Series. D. K. Arora and G. G. Khachatourians (eds.). Elsevier Science, Burlington, MA.

Cesati, V. 1852. Klotzsch, herbarium vivum mycologicum, sistens fungorum per totam Germanium cresentium collection perfectam. Cent. XVII. Flora 35:398.

Chen, F., Goodwin, P. H., Khan, A., and Hsiang, T. 2002. Population structure and mating-type genes of Colletotrichum graminicola from Agrostis palustris. Can. J. Microbiol. 48:427-36.

Couch, B. C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., Kim, P., Notteghem, J.-L., and Kohn, L. M. 2005. Origins of host-specific populations of the rice blast pathogen, *Magnaporthe oryzae*, in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics, in press.

Couch, H. B. 1979. Is it anthracnose or is it wilt? The Greensmaster 15:3-6.

Crouch, J.A., Clarke, B. B., and Hillman, B. I. 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. J., 10: 186-195.

Dale, J. L. 1963. Corn anthracnose. Plant Dis. Rep. 47:245-249.

Darlu, P., and Lecointre, G. 2002. When does the incongruence length difference test fail? Mol. Biol. Evol. 19:432-437.

Dernoeden, P. H. 2000. Stresses and maladies of creeping bentgrass. Pp. 44-48 in: Creeping bentgrass management: Summer stresses, weeds and selected maladies. Ann Arbor Press, Chelsea, MI.

Dolphin, K., Belshaw, R., Orme, C. D. L., and Quicke, D. L. J. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17:401-406.

Dowton, M., and Austin, A. D. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy--the behavior of the incongruence length difference test in mixed-model analysis. Syst. Biol. 51:19-31.

Du, M., Schardl, C. L., Nuckles, E. M., and Vaillancourt, L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Colletotrichum* species complexes. Mycologia, 97(3):641-58.

Fang, G. C., Hanau, R. M., and Vaillancourt, L. J. 2002. The *SOD2* gene, encoding a manganese-type superoxide dismutase, is up-regulated during conidiogenesis in the plant-pathogenic fungus *Colletotrichum graminicola*. Fungal Genet. Biol. 36:155-65.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.

Freeman, S. 2000. Genetic diversity and host specificity of *Colletotrichum* species on various fruits. Pp. 131-144 in: *Colletotrichum*: Host specificity, pathology, and host-pathogen interaction. D. Prusky, S. Freeman and M. B. Dickman (eds.). APS Press, St. Paul, MN.

Hey, J., and Wakeley, J. 1997. A coalescent estimator of the population recombination rate. Genetics 145:833-846.

Horvath, B. J., and Vargas, J. M. 2004. Genetic variation among *Colletotrichum graminicola* isolates from four hosts using isozyme analysis. Plant Dis. 88:402-406.

Hsiang, T., and Goodwin, P. H. 2001. Ribosomal DNA sequence comparisons of *Colletotrichum graminicola* from turfgrasses and other hosts. Eur. J. Plant Pathol. 107:593-599.

Hudson, R. R., and Kaplan, N. L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147-164.

Huelsenbeck, J. P., and Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Huelsenbeck, J. P., and Rannala, B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitutional models. Syst. Biol. 53:904-913.

Huson, D. H. 1998. SplitsTree: Analyzing and visualizing evolutionary data. Bioinformatics 14:68-73.

Jamil, F. F., and Nicholson, R. L. 1987. Susceptibility of corn to isolates of *Colletotrichum graminicola* pathogenic to other grasses. Plant Dis. 71:809-810.

Johnston, P. R., and Jones, D. 1997. Relationships among *Colletotrichum* isolates from fruit-rots assessed using rDNA sequences. Mycologia 89:420-430.

Kemp, G. H. J., Pretorius, Z. A., and Smith, J. 1991. Anthracnose of wheat in South Africa. Phytophylactia 23:177-179.

Larget, B., and Simon, D. L. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750-759.

LeBeau, F. J. 1950. Pathogenicity studies with *Colletotrichum* from different hosts on sorghum and sugarcane. Phytopathology 40:430-438.

LeBeau, F. J., Stokes, I. E., and Coleman, O. H. 1951. Anthracnose and red rot of sorghum. Vol. 1035. U.S. Dept. of Agriculture, Washington, D.C.

Liyanage, H. D., McMillan, R. T., and Kistler, H. C. 1992. Two genetically distinct populations of *Colletotrichum gloeosporioides* from citrus. Phytopathology 82:1371-1376.

Messiaen, C. M., Lafon, R., and Malot, O. 1959. Necroses de racines, pourritures de tiges et verse parasitaire du Mais. Ann. Epiphyt. ser. C. 10:441-474.

Minussi, E., and Kimati, H. 1979. Taxonomy of *Colletotrichum graminicola* (Ces.) Wils. (sensu Arx, 1957). Rev. Centro Cienas Rurais 9:171-187.

Moriwaki, J., Tsukiboshi, T., and Sato, T. 2002. Grouping of *Colletotrichum* species in Japan based on rDNA sequences. J. Gen. Plant Pathol. 68:307-320.

O'Donnell, K., and Cigelnik, E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol. Phylogenet. Evol. 7:103-116.

Posada, D., and Crandall, K. A. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Posada, D., and Crandall, K. A. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends Ecol. Evol. 16:37-45.

Posada, D., and Crandall, K. A. 2002. The effect of recombination on the accuracy of phylogeny estimation. J. Molec. Evol. 54:396-402.

Posada, D., Crandall, K. A., and Holmes, E. C. 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36:75-97.

Price, E. W., and Carbone, I. 2005. SNAP: Workbench management toll for evolutionary population genetic analysis. Bioinformatics 21:402-404.

Randhir, R. J., and Hanau, R. M. 1997. Size and complexity of the nuclear genome of *Colletotrichum graminicola*. Appl. Envir. Microbiol. 63:4001-4004.

Rozas, J., Sanchez-DelBarrio, J. C., Messeguel, X., and Rozas, R. 2003. DnaSP, DNA sequence polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Selby, A. D., and Manns, T. F. 1909. Studies in diseases of cereals and grasses. Ohio Agr. Expt. Sta. Bul. 203:207.

Sherriff, C., Whelan, M. J., Arnold, G. M., and Bailey, J. A. 1995. rDNA sequence analysis confirms the distinction between *Colletotrichum graminicola* and *C. sublineolum*. Mycol. Res. 99:475-478.

Sreenivasaprasad, S., Mills, P. R., Meehan, B. M., and Brown, A. E. 1996. Phylogeny and systematics of 18 *Colletotrichum* species based on ribosomal DNA spacer sequences. Genome 39:499-512.

Sullivan, J. 1996. Combining data with different distributions of among-site rate variation. Syst. Biol. 45:375-380.

Sutton, B. C. 1965. Studies on the taxonomy of *Colletotrichum* Cda with especial reference to *C. graminicola* (Ces.) Wilson, University of London, London.

Sutton, B. C. 1966. Development of fruitifications in *Colletotrichum graminicola* (Ces.) Wils. and related species. Can. J. Bot. 44:887-897.

Sutton, B. C. 1968. The appressoria of *Colletotrichum graminicola* and *C. falcatum*. Can. J. Bot. 46:873-876.

Sutton, B. C. 1980. The coelomycetes: fungi imperfecti with pycnidia, acervuli, and stromata. Kew, U.K.: Commonwealth Mycological Institute.

Sutton, B. C. 1992. The genus *Glomerella* and its anamorph *Colletotrichum*. Pp. 1-26 *In Colletotrichum*: Biology, pathology and control. J. A. Bailey and M. J. Jeger (eds.). CAB International, Wallingford, U.K.

Swofford, D. L. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S., and Fisher, M. C. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31:21-32.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. Nucleic Acids Res. 22:4673-4680.

Vaillancourt, L., Du, M., Wang, J., Rollins, J., and Hanau, R. 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia 92:430-435.

Vaillancourt, L. J., and Hanau, R. M. 1991. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. Phytopathology 81:530-534.

Vaillancourt, L. J., and Hanau, R. M. 1992. Genetic and morphological comparisons of *Glomerella* (*Colletotrichum*) isolates from maize and sorghum. Exp. Mycol. 16:219-229.

Valero, H. M., Resende, M. A., Weikert-Oliveira, R. C. B., and Casela, C. R. 2005. Virulence and molecular diversity in *Colletotrichum graminicola* from Brazil. Mycopathologia 159:449-459.

White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 *In* PCR Protocols: A Guide to Methods and Applications. M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White (eds.). Academic Press, San Diego.

Wilson, G. W. 1914. The identity of the anthracnose of grasses in the United States. Phytopathology 4:106-112.

Zeiders, K. E. 1987. Leaf spot of Indiangrass caused by *Colletotrichum caudatum*. Plant Dis. 71:348-350.

Zwillenberg, H. H. L. 1959. *Colletotrichum graminicola* (Ces.) Wilson auf mais und verschiedenen anderen Pflanzen. Phytopath. Z. 34:417-425.

Table 1.1

Sources of *Colletotrichum* isolates used in this study.

Clade	Haplotype (frequency)	Isolate Name	Host Species	Origin	Source	Original Reference
--	1 (1)	NJ-CACA (outgroup)	*Calamagrostis acutifolia* (*C. acutatum*)	New Jersey	--	This study
B	2 (1)	ALB-99325	*Poa pratensis*	Alberta, Canada	T.Hsiang	This study
A	3 (1)	MA-6722	*Poa annua*	Massachusetts	--	This study
A	4 (1)	CT-6956	*Poa annua*	Connecticut	--	This study
A	5 (22)	CA-62	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-1049	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-1143	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-CL9	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-SH29	*Poa annua*	California	F. Wong	This study
A	5 (22)	CT-2	*Poa annua*	Connecticut	N. Jackson	(12)
A	5 (22)	IL-P6G	*Poa annua*	Illinois	N. Jackson	(12)
A	5 (22)	IL-PT	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	IL-PV1	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	IL-PV2	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	IL-RCC	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	NJ-6795	*Poa annua*	New Jersey	--	This study
A	5 (22)	NJ-7284	*Poa annua*	New Jersey	--	This study
A	5 (22)	NY-8422	*Poa annua*	New York	--	This study
A	5 (22)	NY-USGA	*Poa annua*	New York	--	This study
A	5 (22)	ONT-00176	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
A	5 (22)	PA-50014	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-50111	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-50231	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-V1	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-WH3	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	RI-8	*Poa annua*	Rhode Island	N. Jackson	(12)
A	6 (1)	NBR-13	*Poa annua*	New Brunswick, Canada	N. Jackson	(12)
A	7 (1)	CA-540	*Poa annua*	California	F. Wong	This study
A	8 (1)	CO-8910	*Poa annua*	Colorado	--	This study
A	9 (4)	KS-20DGU	*Dactylis glomerata*	Kansas	--	This study
A	9 (4)	KS-20DGY	*Dactylis glomerata*	Kansas	--	This study
A	9 (4)	KS-20EVD	*Elymus virginicanus*	Kansas	--	This study

A	9 (4)	KS-20EVM	*Elymus virginicanus*	Kansas	--	This study
A	10 (1)	NJ-CA1	*Calamagrostis acutifolia*	New Jersey	--	This study
A	11 (1)	NJ-DG1	*Dactylis glomerata*	New Jersey	--	This study
A	12 (1)	NJ-8627	*Poa annua*	New Jersey	--	This study
A	13 (1)	KS-10EC1A	*Elymus canadensis*	Kansas	--	This study
A	14(1)	NY-8900	*Poa annua*	New York	--	This study
A	15 (1)	NJ-8626	*Poa annua*	New Jersey	--	This study
A	16 (1)	NJ-6340	*Poa annua*	New Jersey	--	This study
A	17 (2)	CA-EG15	*Poa annua*	California	F. Wong	This study
A	17 (2)	CA-SC44	*Poa annua*	California	F. Wong	This study
A	18 (1)	MA-11	*Poa annua*	Massachusetts	N. Jackson	(12)
A	19 (1)	CT-14	*Poa annua*	Connecticut	N. Jackson	(12)
A	20 (1)	KS-20BIG	*Bromus inermis*	Kansas	--	This study
A	17 (2)	NJ-9582	*Poa annua*	New Jersey	--	This study
A	17 (2)	PA-50183	*Poa annua*	Pennsylvania	W. Uddin	This study
A	22 (1)	CA-SC32	*Poa annua*	California	F. Wong	This study
A	23 (1)	CA-1715	*Poa annua*	California	F. Wong	This study
A	24 (1)	RI-22	*Agrostis stolonifera*	Rhode Island	N. Jackson	(12)
B	25 (1)	PA-50002	*Poa annua*	Pennsylvania	W. Uddin	This study
B	26 (1)	CT-28	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
B	27 (1)	NJ-6607	*Poa annua*	New Jersey	--	This study
B	28 (1)	MA-17	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	29 (1)	ONT-00133	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
B	30 (1)	CT-25	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
B	31 (3)	MA-20	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	31 (3)	MA-21	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	31 (3)	ONT-00128	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study

Table 1.1

Sources of *Colletotrichum* isolates used in this study, continued.

Clade	Haplotype (frequency)	Isolate Name	Host Species	Origin	Source	Original Reference
B	32 (2)	ONT-00124	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	(16)
B	32 (2)	ONT-00126	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
B	33 (1)	PA-50621	*Poa annua*	Pennsylvania	W. Uddin	This study
B	34 (1)	NY-19	*Agrostis stolonifera*	New York	N. Jackson	(12)
B	35 (1)	PA-4410	*Poa annua*	Pennsylvania	W. Uddin	This study
B	36 (1)	NJ-6491	*Poa annua*	New Jersey	--	This study
B	37 (2)	NJ-4990	*Poa annua*	New Jersey	--	This study
B	37 (2)	PA-1112	*Poa annua*	Pennsylvania	W. Uddin	This study
B	38 (2)	PA-50005	*Poa annua*	Pennsylvania	W. Uddin	This study
B	38 (2)	PA-50181	*Poa annua*	Pennsylvania	W. Uddin	This study
B	39 (1)	MA-24	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	40 (1)	PA-50623	*Poa annua*	Pennsylvania	W. Uddin	This study
B	41 (1)	CT-18	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
A	42 (1)	TX-26	*Agrostis stolonifera*	Texas	N. Jackson	(12)
A	43 (1)	VA-8977	*Poa annua*	Virginia	--	This study
A	44 (1)	ONT-99359	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
A	45 (1)	NH-23	*Agrostis stolonifera*	New Hampshire	N. Jackson	(12)
A	46 (1)	NJ-8467	*Poa annua*	New Jersey	--	This study
A	47 (5)	NJ-7130	*Agrostis stolonifera*	New Jersey	--	This study
A	47 (5)	NJ-8400	*Lolium perenne*	New Jersey	--	This study
A	47 (5)	NJ-RWCC	*Poa annua*	New Jersey	--	This study
A	47 (5)	NJ-HF2B	*Agrostis stolonifera*	New Jersey	--	This study
A	47 (5)	PA-211	*Poa annua*	Pennsylvania	W. Uddin	This study
B	48 (1)	NJ-6553	*Poa annua*	New Jersey	--	This study
B	49 (1)	CT-27	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
B	50 (1)	NY-16	*Agrostis stolonifera*	New York	N. Jackson	This study
--	51 (1)	NJ-10BB (outgroup)	*Vaccinium corymbosum* (*C. acutatum*)	New Jersey	P. Oudemans	This study
C	52 (1)	S3001	*Sorghum bicolor* (*C. sublineolum*)	Burkina Fasso	L. Vaillancourt	(67)
D	53 (1)	NY-15182	*Zea mays*	New York	G. Bergstrom	This study
C	54 (2)	S12001	*Sorghum bicolor*	Brazil	L. Vaillancourt	(67)
C	54 (2)	S19001	*Sorghum bicolor*	South Africa	L. Vaillancourt	(25)

C	55 (1)	S17001	*Sorghum bicolor*	Texas	L. Vaillancourt	(25)
D	56 (2)	MO-478	*Zea mays*	Missouri	L. Vaillancourt	This study
D	56 (2)	MO-978	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	BZ-500190	*Zea mays*	Brazil	L. Vaillancourt	(67)
D	57 (18)	IN-10472	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-10570	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-10670	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-10970	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-12270	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-12475	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-300170	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-DUB90	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	KY-197	*Zea mays*	Kentucky	L. Vaillancourt	This study
D	57 (18)	KY-297	*Zea mays*	Kentucky	L. Vaillancourt	This study
D	57 (18)	KY-397	*Zea mays*	Kentucky	L. Vaillancourt	This study
D	57 (18)	KY-398	*Zea mays*	Kentucky	L. Vaillancourt	(25)
D	57 (18)	MO-178	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	MO-878	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	MO-1001178	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	NC-200170	*Zea mays*	North Carolina	L. Vaillancourt	This study
D	57 (18)	NY-AU80	*Zea mays*	New York	L. Vaillancourt	(25)
D	57 (18)	NY-400180	*Zea mays*	New York	L. Vaillancourt	This study
D	58 (1)	IN-D77	*Zea mays*	Indiana	L. Vaillancourt	This study
D	59 (1)	IN-900190	*Zea mays*	Indiana	L. Vaillancourt	This study

TABLE 1.2: Characteristics of the genomic regions used in this investigation.

Gene	Exon regions	Intron regions	Nucleotide characters, excluding indels	Coded characters (indels)	Invariable nucleotide characters*	Variable, parsimony uninformative nucleotide characters*	Parsimony informative nucleotide characters*	MP tree length	Number of equally parsimonious trees
ITS	1	2	482	49	358	68	56 *(12%)*	357	44,034
HMG	2	1	211	13	170	14	26 *(12%)*	139	4
SOD	3	2	536	38	504	0	221 *(41%)*	518	72
Combined Dataset	6	5	1229	130	1032	82	303 *(25%)*	1058	8,813

* Parsimony informative characterization of dataset considers ingroup taxa only, outgroups are excluded

TABLE 1.3: Best fit evolutionary models used in Bayesian analysis as determined by the AIC in Modeltest.

ITS 1 / 5.8S / ITS 2 **482 nucleotide characters**	**HMG (*Mat 1-2*) 211 nucleotide characters**	***Sod-2*** **536 nucleotide characters**
SYM+G	**HKY+G**	**HKY+G**
A→C 1.0161	Frequency A = 0.305	Frequency A = 0.2352
A→G 1.2515	Frequency C = 0.269	Frequency C = 0.2575
A→T 1.0096	Frequency G = 0.224	Frequency G = 0.3088
C→G 1.7520	Frequency T = 0.201	Frequency T = 0.1986
C→T 2.7189		
G→T 1.000		
	Ti/Tv ratio = 1.2350	Ti/Tv ratio = 1.5889
Invariable = 0	Invariable = 0	Invariable = 0
Gamma	Gamma	Gamma
α = 0.9962	α = 1.8284	α =0.8737

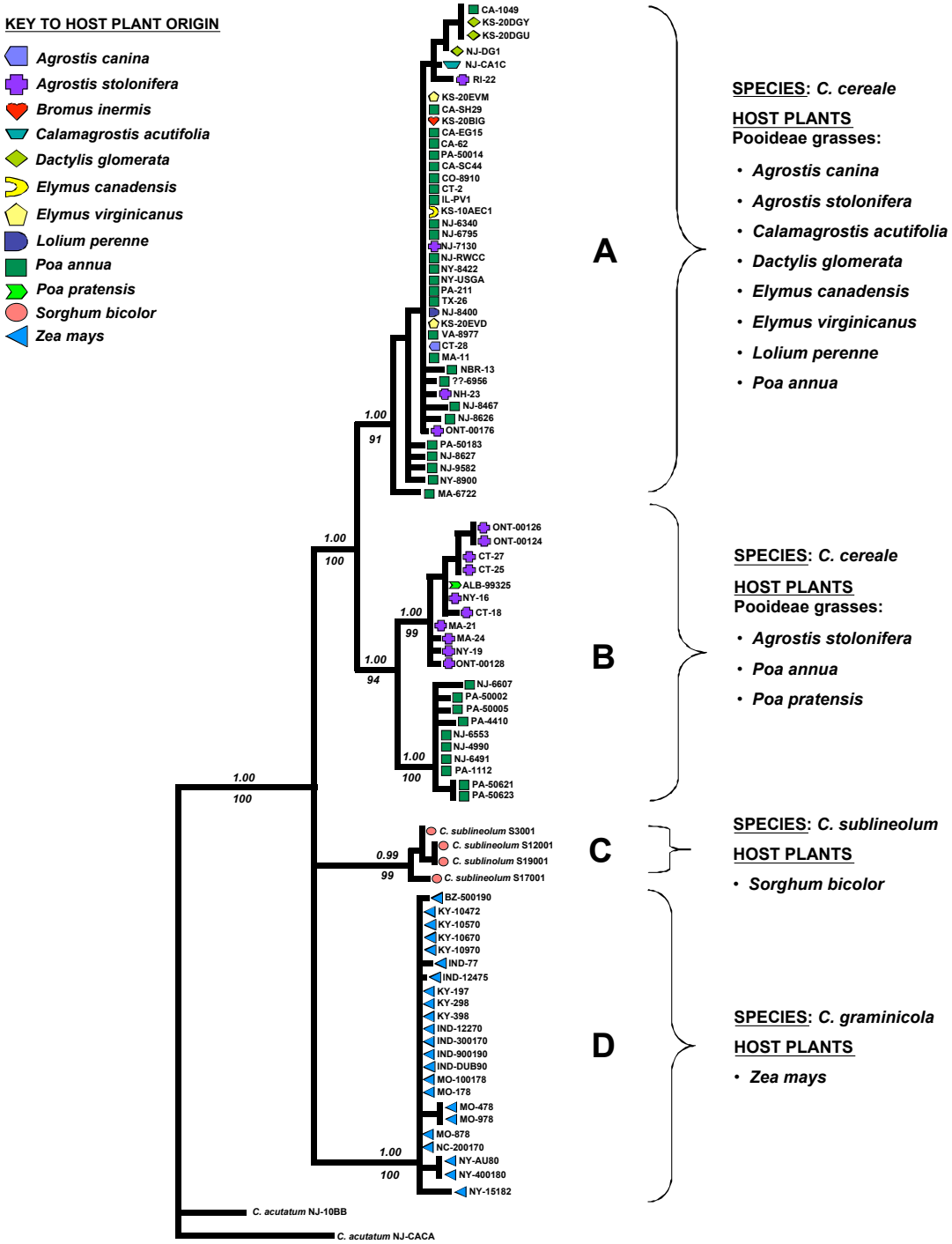


Figure 1.1

75% majority rule Bayesian ML phylogenetic tree. All groups illustrated by this tree are supported by posterior probabilities of at least 0.75. Posterior probabilities supporting the topology are shown above the nodes where space permits; elsewhere on the tree two asterisks were supported by posterior probabilities of >99%; branches with one asterisk were supported by posterior probabilities of >95%. MP bootstrap values from 1,000 pseudoreplicates are shown below the nodes.

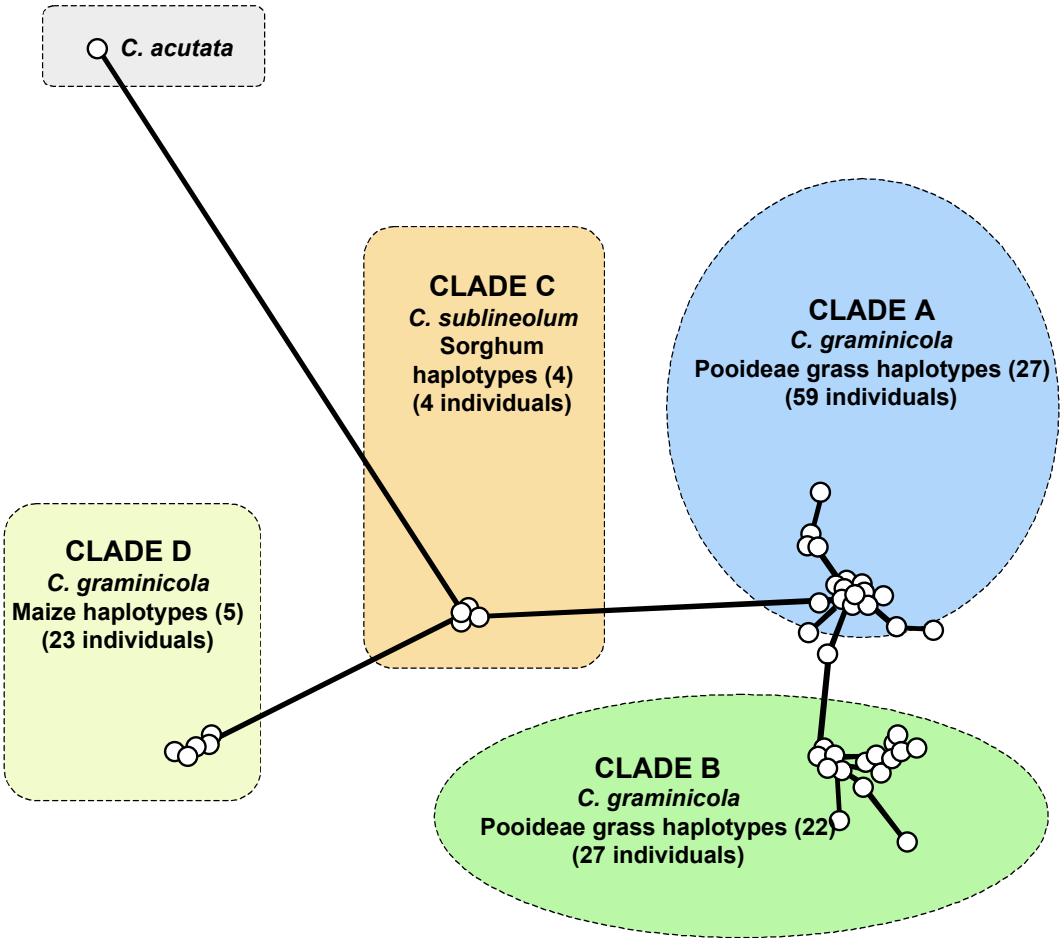


Figure 1.2
Evolutionary relationships, as inferred using split-decomposition network analysis. Four main groups were recovered; these groups correspond to the phylogenetically-based lineages A--D. No reticulate evolution is evident between clades, but cycles connect many of the haplotypes within each individual clade. The splits-graph suggests that *C. graminicola* and *C. cereale* are each significantly more closely related to *C. sublineolum* than they are to one another.

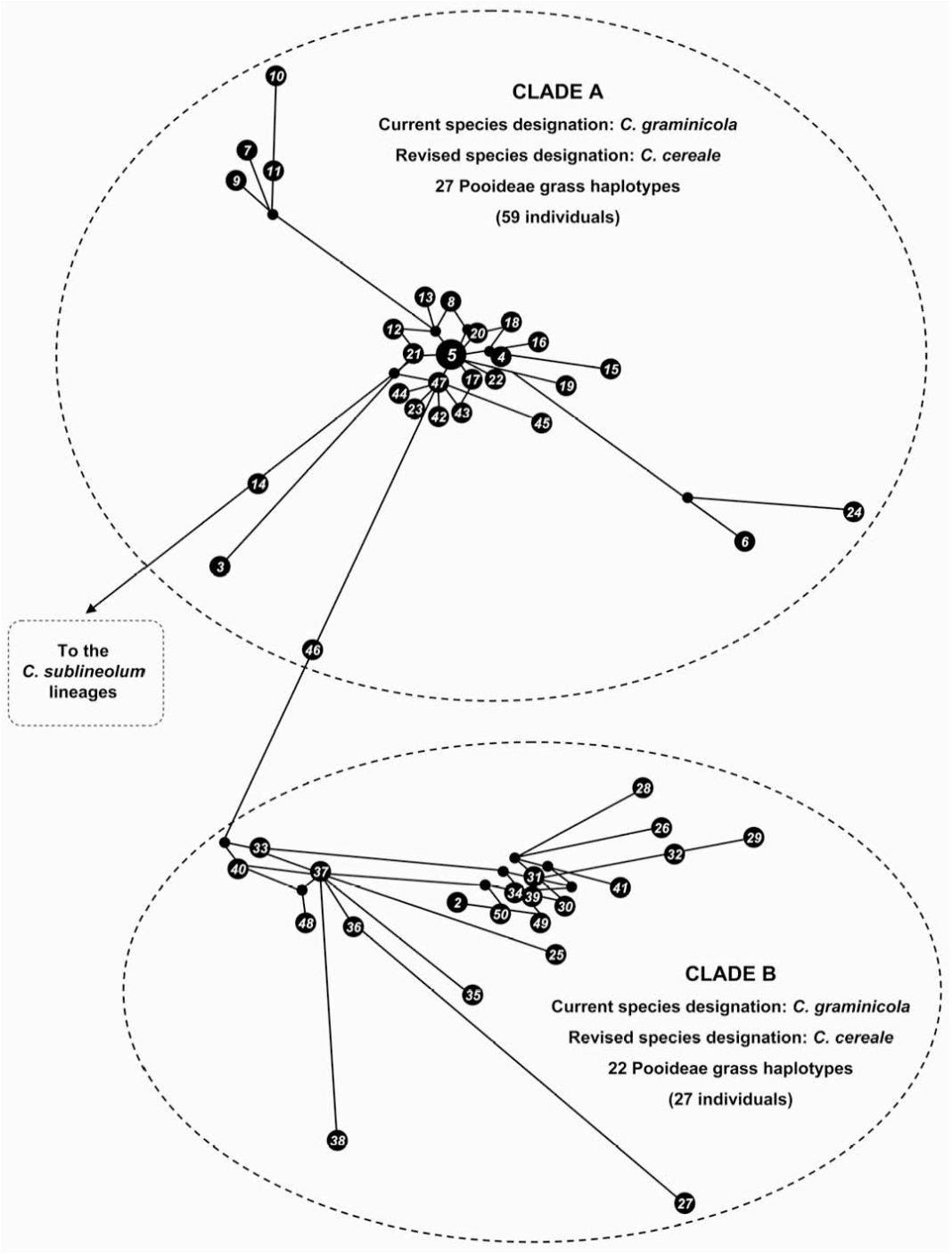


Figure 1.3
Enlarged view of the splits-graph topology, illustrating *C. cereale* clades A and B. The small filled circles represent the positions of inferred intermediate haplotypes. Both of the *C. cereale* clades are characterized by highly reticulate relationships, but appear to be distinct from one another.

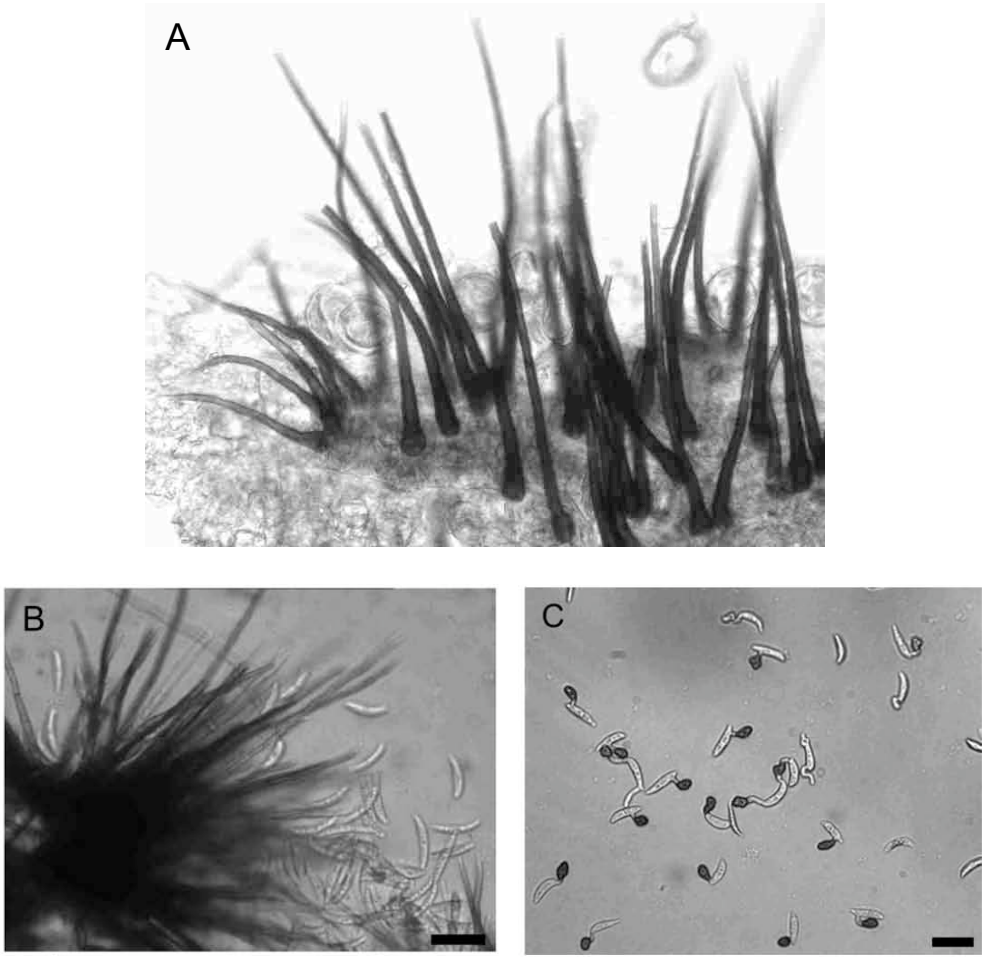


Figure 1.4
C. cereale morphological structures (Bar=25 μm). (A) Heavily melanized setae emerging from an acervulus on the leaf of *Bromus secalinus* from the lectotype specimen (NYBG-305598); (B) setae and conidia, grown on PDA from epitype strain KS-20BIG; and (C) germinating conidia, with the formation of germ tubes generally (but not always) leading to the formation of heavily melanized appressoria.

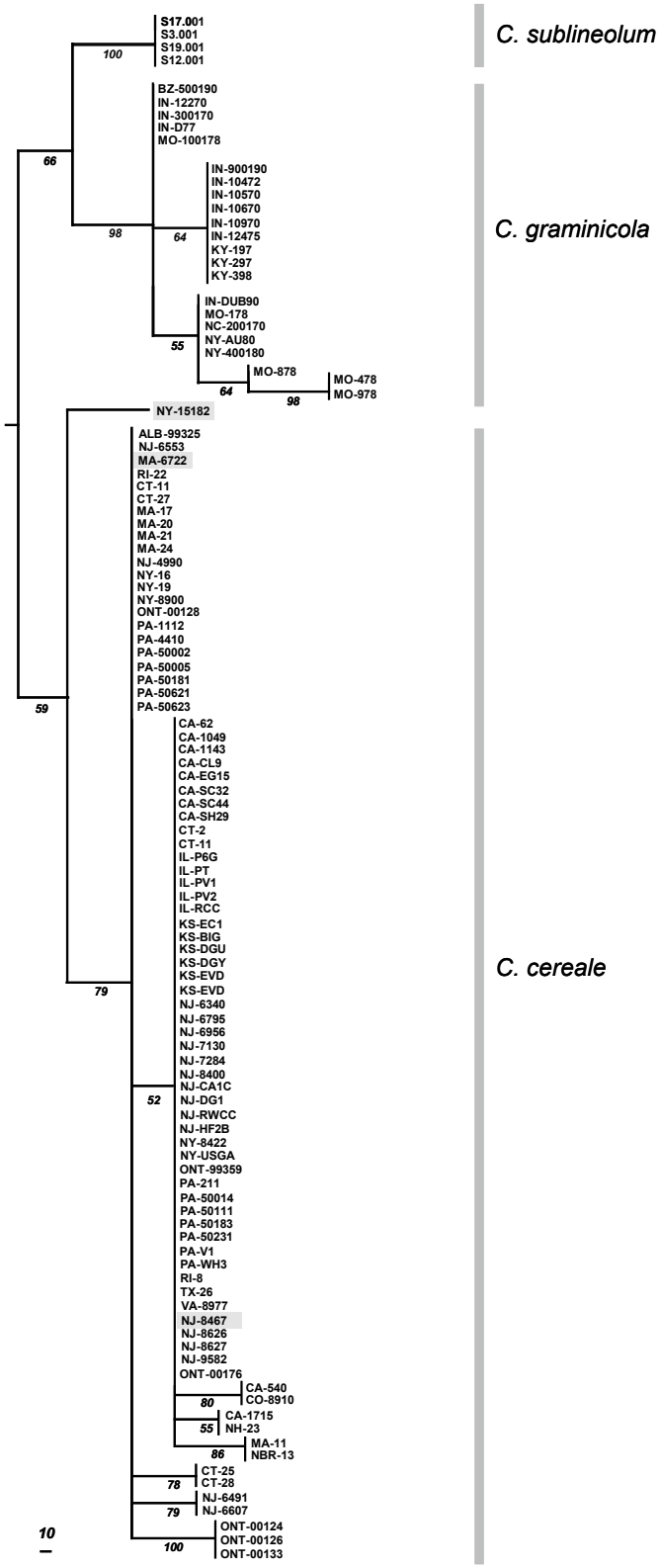


Figure 1.5
ITS gene maximum parsimony phylogenetic tree.

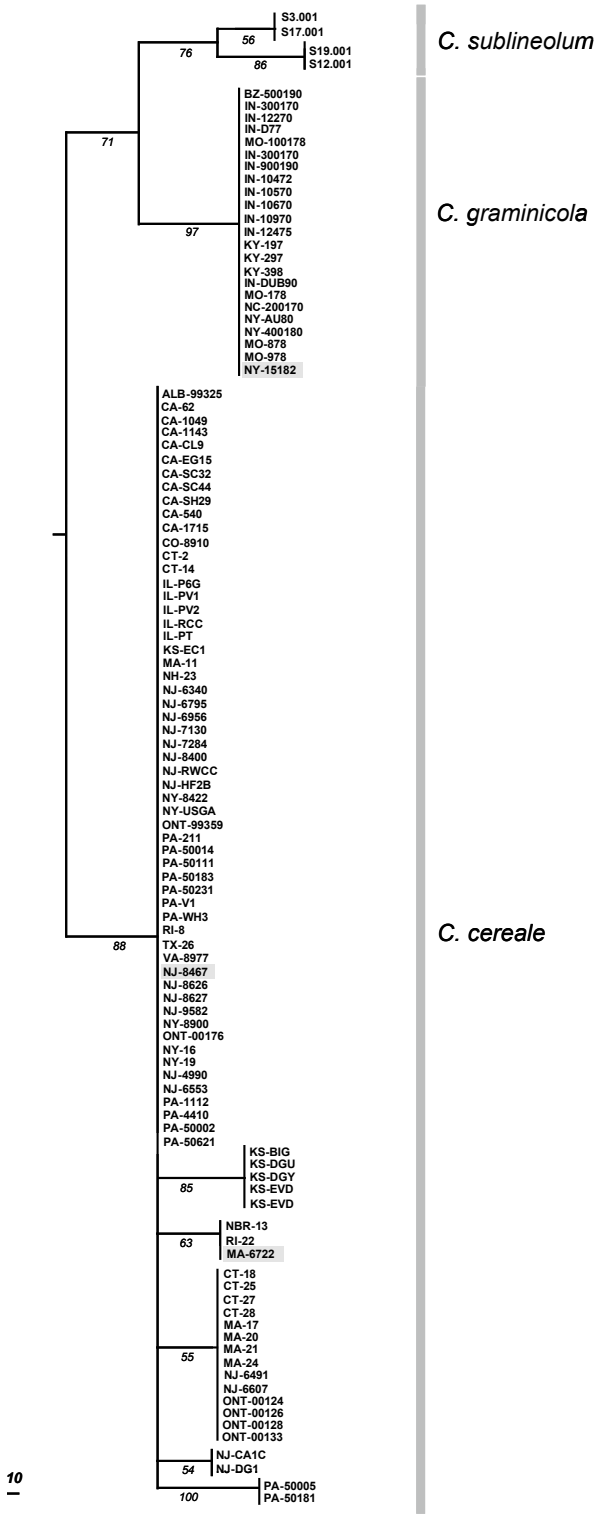


Figure 1.6.
HMG gene maximum parsimony phylogenetic tree.

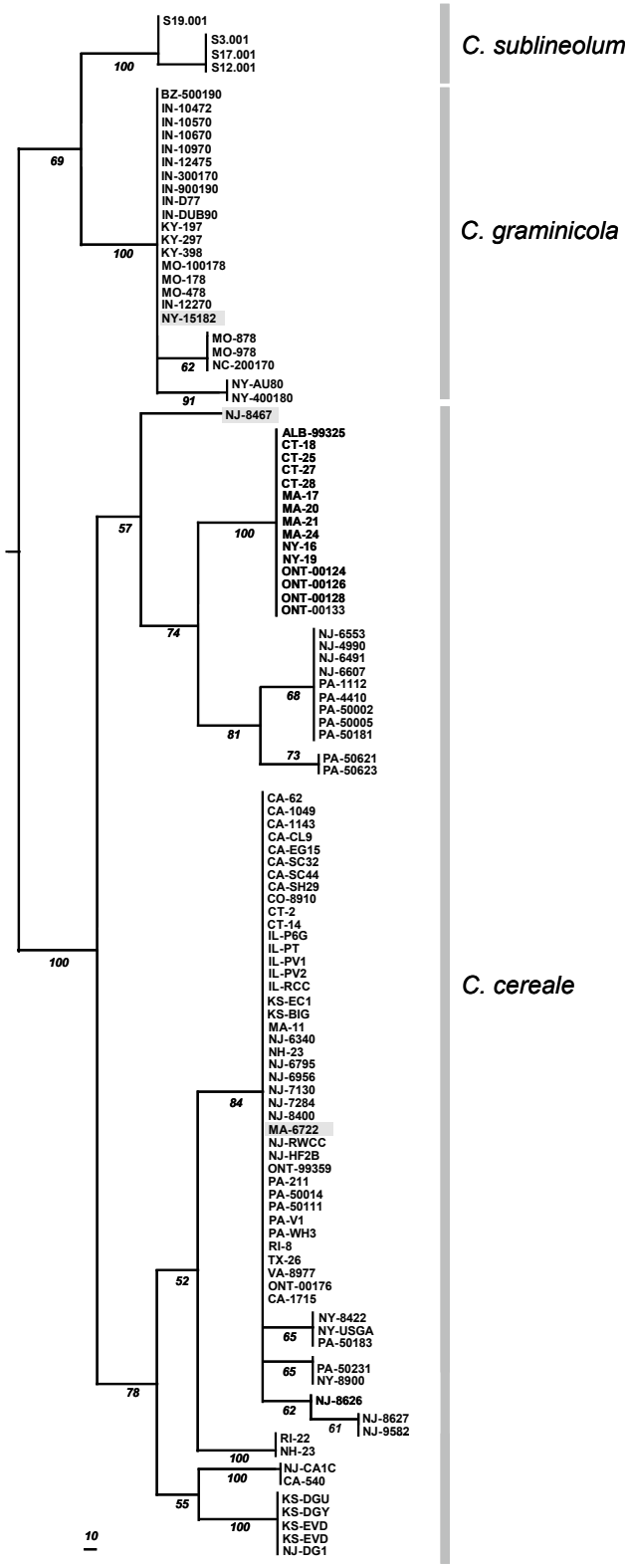


Figure 1.7
Sod2 gene maximum parsimony phylogenetic tree.

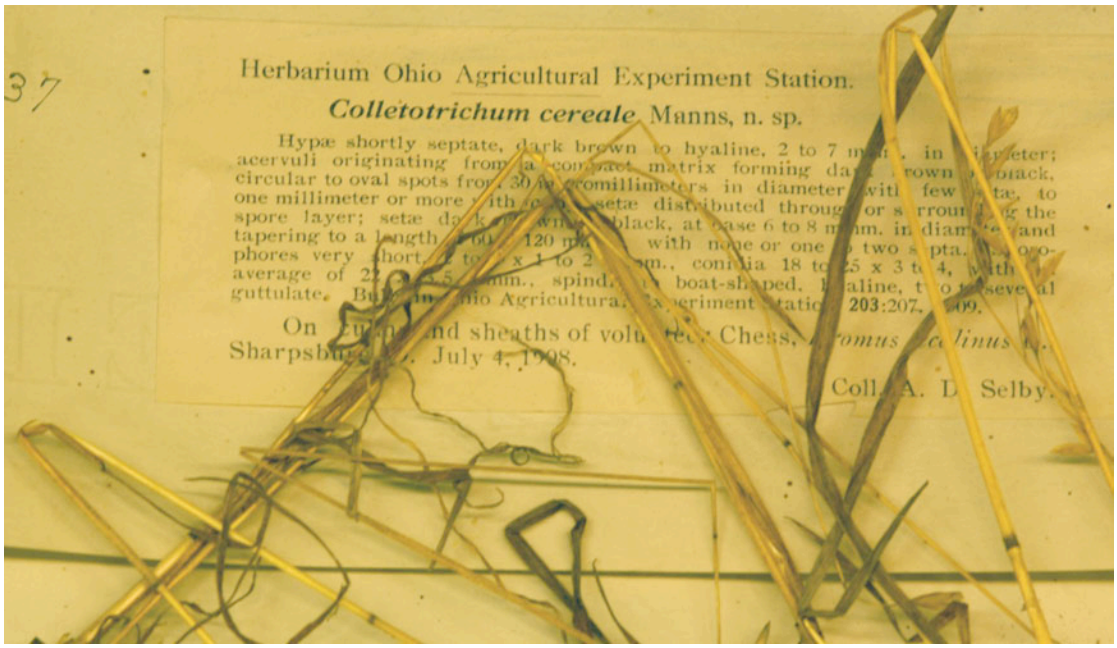


Figure 1.8
Photograph of original *Colletotrichum cereale* lectotype specimen from the Kansas State Herbarium, presently curated by the New York Botanical Garden Steere Herbarium (NYBG).) NYBG specimen numbers 305598 on *Bromus secalinus*..

Chapter 2: The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen

ABSTRACT

Mobile transposable elements are among the primary drivers of the evolution of eukaryotic genomes. For fungi, repeat-induced point mutation (RIP) silencing minimizes deleterious effects of transposons by mutating multicopy DNA during meiosis. In this study we identify five transposon species from the mitosporic fungus *Colletotrichum cereale* and report the signature pattern of RIP acting in a lineage-specific manner on 21 of 35 unique transposon copies, providing the first evidence for sexual recombination for this species. Sequence analysis of genomic populations of the retrotransposon *Ccret2* shows that repeated rounds of RIP mutation have acted on different copies of the element. In a population of RIPped *Ccret2* elements, multiple inferences of incongruence were attributed primarily to RIP-induced homoplasy. This study supports the view that the sequence variability of transposon populations in filamentous fungi reflects the activities of evolutionary processes that fall outside of typical phylogenetic or population genetic reconstructions.

2.1 Introduction

Mobile genetic elements such as transposons (TEs) are abundant in eukaryotes, and with the exception of *Plasmodium falciparum* (Gardner *et al.*, 2002), the causative agent

of human malaria, TEs populate the DNA of all well-studied organisms. TEs may occupy a substantial proportion of the host genome: 60% of the maize genome is transposon-derived (Messing and Dooner, 2006), as is 38% of the mouse genome (IMGSC, 2001) and 45% of the human genome (IHGSC, 2001). In contrast, the genomes of many eukaryotes are composed of relatively few transposons: for example, only 4.3% of the chicken genome is transposon-derived (Wicker *et al.*, 2005). A relatively small contribution of TEs to the genomes of fungi is typical, with only 3.1% of the *Saccharomyces cereviseae* genome comprised of TEs (Goffeau *et al.*, 1996), while 8.2%-14% of the genome of the rice blast fungus, *Magnaporthe oryzae* may be derived from TEs (Dean *et al.*, 2005, Thon *et al.*, 2006).

Because TEs are able to move about the host genome and insert into a host's DNA through either cut-and-paste (DNA, or Class II transposons) or copy-and-paste mechanisms via RNA intermediates (retro, or Class I transposons), these elements can exert a significant influence on the fitness and evolutionary potential of their hosts through events such as insertional mutagenesis, disrupted or enhanced gene expression or gross chromosomal rearrangements (Hua-Van *et al.*, 2005). Given the numerous ways that transposition can impact the genome, a variety of methods have evolved to safeguard the host against the effects of potentially deleterious insertions or unsupportable transposition rates. In several organisms, highly specific targeting mechanisms have been shown to limit TE integration to non-essential genomic regions, thereby protecting host integrity. Most TEs appear to have integration "hotspots" that are dictated by nucleotide sequence, patterns of hydrogen bonds, DNA-bending proteins and/or DNA conformation (Chalmers *et al.*, 1998, Bender and Kleckner, 1992, Ketting *et al.*, 1997, Liu *et al.*, 2005). Well known examples of targeted integration in fungi include the Ty

retroelements of *S. cerevisiae*, which insert preferentially upstream of pol III transcribed genes and in silent chromatin regions (Zou *et al.*, 1996, Devine and Boeke, 1996, Chalker and Sandmeyer, 1992) and the retrotransposon Tf1 in *Schizosaccharomyces pombe*, which exhibits a clear preference for integration in tandem and divergent intergenic pol II promoter regions (Singleton and Levin, 2002).

Filamentous fungi actively regulate repetitive sequences through silencing mechanisms such as quelling (RNA silencing), (Cogoni *et al.*, 1996), meiotic silencing (Shiu *et al.*, 2001), and repeat-induced point mutation (RIP) (Cambareri *et al.*, 1989). The RIP mutation process is remarkably efficient in disabling transposable elements through the detection and subsequent mutation of duplicated sequences longer than ~400-bp (Watters *et al.*, 1999). Just prior to karyogamy, GC-to-AT transitions are induced in duplicate sequences sharing >80% similarity, with as many as 30% of GCs converted to ATs (Cambareri *et al.*, 1989) and repetitive DNA remaining susceptible to "RIPping" through six generations (Cambareri *et al.*, 1991). Since its initial discovery in *Neurospora crassa* (Selker *et al.*, 1987), the RIP-mutation process has been identified experimentally or through sequence analysis in the ascomycetes *Aspergillus fumigatus* (Neuveglise *et al.*, 1996), *A. nidulans* (Nielson *et al.*, 2001, Clutterbuck, 2004), *A. oryzae* (Montiel *et al.*, 2006) *Fusarium oxysporum* (Hua-Van *et al.*, 2001), *Leptosphaeria maculans* (Attard *et al.*, 2005), *M. oryzae* (Nakayashiki *et al.*, 1999, Ikeda *et al.*, 2002), *N. tetrasperma* (Bhat *et al.*, 2004); *Ophiostoma* sp. (Bouvet *et al.*, 2007) and *Podospora anserina* (Graia *et al.*, 2001) and in the basidiomycete *Microbotryum violacum* (Hood *et al.*, 2005), although RIP-mutation activity in these species has always been found in a much less aggressive form than that observed in *N. crassa* (Galagan and Selker, 2004).

We have been developing several molecular tools, including transposon-based

marker systems, to increase our understanding of the recent emergence of the mitosporic Ascomycete fungus *Colletotrichum cereale* as a pathogen of turfgrasses and its benign existence in natural grassland and agroecosystems. Beginning in the mid 1990s, *C. cereale* emerged as one of the most destructive pathogens of cool-season turfgrasses (Smiley *et al.*, 2005). Outside of golf course greens, *C. cereale* is a common inhabitant of a wide range of C3 cereals and grasses of the grass subfamily *Pooideae*, where it survives without inducing noticeable levels of disease (Crouch *et al.*, 2006, J.A. Crouch and B.I. Hillman, *unpublished data*). Little is known about *C. cereale* populations and, until recently, the fungus was thought to be conspecific with *C. graminicola*, a pathogen of corn (Crouch *et al.*, 2006). Two major *C. cereale* lineages (clades A and B) have been described using sequences of the internal transcribed spacer (ITS) region of the ribosomal DNA (Crouch *et al.*, 2005) and multilocus phylogenetic analyses (Crouch *et al.*, 2006), but the evolutionary processes that shaped these lineages remain largely unexplored.

During the course of surveying the *C. cereale* genome for TEs suitable for use as molecular markers, we observed that many of this organism's transposon sequences were distinguished by a pronounced A+T nucleotide bias; subsequent bioinformatics analysis demonstrated this bias reflected the characteristic patterns of RIP-like C→T and G→A transitions. In this study, five different species of transposons were identified from the two major lineages of *C. cereale* in RIPped and "normal", non-mutated forms: two DNA transposons, two species of long terminal repeat (LTR) retrotransposons and one non-LTR retrotransposon. In this paper we describe these *C. cereale* transposable elements and document how the process of RIP mutation has considerably altered 21 of 35 unique transposon copies surveyed in a lineage specific manner. We then employ the *Ccret2* retrotransposon *pol* gene sequence to explore the impact of RIP mutated transposons

when these elements are used to generate evolutionary inferences for phylogenetic and population genetic analyses.

2.2 Materials and methods

2.2.1 Construction of genomic DNA libraries and the identification of repetitive transposon DNA

Genomic DNA was isolated from fungal mycelia using a standard phenol:chloroform extraction protocol (Sambrook and Russell, 2001). *C. cereale* genomic DNA libraries were constructed from the *Eco*RI-digested DNA of isolates PA-50231 (clade A) and PA-50005 (clade B) in the plasmid vector pGEM3zf+ (Promega, Madison, WI); the culture and origin of these isolates was described previously (Crouch *et al.*, 2006). To screen for repetitive sequences, insert-bearing colonies were transferred to Colony/Plaque Screen Transfer Membranes (NEN, Boston, MA) following the manufacturer's alkaline lysis protocol. The Random Primers DNA Labeling System (Invitrogen Corp., Carlsbad, CA) was used to label *C. cereale* genomic DNA with [α^{32}P]dCTP (MP Biomedicals, Irvine, CA). After a 30 min prehybridization in hybridization solution [7% SDS, 1 mM EDTA and 0.5 M Na_2HPO_4] at 65°C, the membranes were hybridized overnight with the denatured, labeled total genomic DNA probe in fresh hybridization solution at 65°C. Two high stringency washes were performed at 65°C [5% SDS, 1 mM EDTA and 40 mM Na_2HPO_4] for 20 min per wash; followed by two additional washes [1% SDS, 1 mM EDTA and 40 mM Na_2HPO_4] for 20 min each. The hybridized membranes were exposed to autoradiography film (Lab

Scientific, Inc., Livingston, NJ) in the presence of an intensifying screen and incubated at -70°C for 24 to 48 hours before development.

Plasmid DNA from clones containing putative repetitive DNA inserts, as predicted by the occurrence of significant levels of hybridization, was purified using the QIAprep Spin Miniprep Kit (Qiagen, Inc., Valencia, CA).

To explore the potential distribution of RIPped TEs in *C. cereale* clade A, an ordered cosmid library of 6,144 clones was constructed in the pWEB vector (EPICENTRE Biotechnologies, Madison, WI) from the genomic DNA of the *C. cereale* species epitype strain NJ-6340 (clade A) (Crouch *et al.*, 2006). Colonies were transferred to Zeta-probe membranes (Bio-Rad Laboratories, Hercules, CA) using the alkaline lysis protocol and probed with [α^{32}P]dCTP-labelled amplicon of the RIP-mutated *C. cereale* retrotransposon *Ccret1*^{DBP6} as described above. High quality plasmid purifications were prepared from the strongly hybridizing cosmid clone 9F8 using the Nucleobond AX Plasmid Maxi kit (BD Bioscience, Easton, PA).

2.2.2 Sequence analysis

Insert DNA was sequenced in both directions from the T7 and SP6 primer sites using the ABI Prism BigDye sequencing chemistry (Applied Biosystems, Inc., Foster City, CA) on an ABI 3100 capillary sequencer. Sequence data was used to perform BLAST searches (Altschul *et al.*, 1990) against the National Center for Biotechnology Information (NCBI) GenBank database to identify sequences with similarity to transposable elements.

Multiple sequence alignments of the putative *C. cereale* transposons and similar transposons identified from GenBank for phylogenetic analyses were generated using

Clustal W (Thompson *et al.*, 1994) as launched in MegAlign (DNASTAR, Inc., Madison, WI), manually adjusted to exclude gaps and ambiguously aligned regions from the dataset. Amino acid alignments were evaluated using maximum likelihood (ML) analyses run in Phylip v3.66 (Felsenstein, 2006) modeled under the best fit evolutionary parameters obtained using ProtTest v1.3 (Abascal *et al.*, 2005). Posterior probabilities supporting the ML tree topologies were generated using Mr. Bayes v3.1 (Huelsenbeck and Ronquist, 2003) by performing two simultaneous runs of one cold and three heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC) for 10,000,000 generations and sampling trees every 500 generations. Fixed-rate evolutionary models for the amino acid data were estimated by MCMCMC model jumping between nine models of protein evolution; each model contributed to the posterior distribution according to the proportion of its posterior probability. Run diagnostics were performed every 1000th generation, with the first 25% of the output discarded as burn-in. To confirm that the two separate runs converged to a stationary distribution, the average standard deviation of split frequencies, the potential scale reduction factor convergence diagnostic, and the plot of generation versus log likelihood values were each examined for values and distributions characteristic of samples drawn from a posterior probability distribution. Trees sampled from the posterior distribution were imported into PAUP* v.4.0b10 (Swofford, 2000) and used to construct 50% majority-rule consensus trees from which support values were derived.

2.2.3 Estimates of repeat-induced point mutation

To evaluate patterns of repeat-induced point mutation in the transposons identified in this study, 110 DNA sequences (GenBank Accession Nos.

DQ663091–DQ663135, DQ666147–DQ666164 and DQ667983–DQ668029) from the *C. cereale* genome were generated from several sources: random sequencing using the universal priming sites of clones from the PA-50005 and PA-50231 plasmid DNA libraries constructed for colony hybridizations (*section 4. 2.1*); sequencing from universal priming sites using a plasmid DNA library constructed from the isolate NJ-6340 in the vector pGEM-T Easy (Promega, Madison, WI); sequencing from universal priming sites using a cDNA library constructed in the vector pGEM-T Easy from *C. cereale* clade A isolate KS-20B-DGU; a 5-kb sequence from the mating type locus from the NJ-6340 cosmid library (*section 4.2.1*); and direct sequencing from the PCR amplicons of isolate NJ-6340 genomic DNA from three single-copy genes (manganese superoxide dismutase, β-tubulin 2, and a class V chitin synthase; Table 1). After removing all ribosomal, mitochondrial and repetitive DNA sequences, a concatenated sequence of 70,594-bp was constructed from the 110 individual sequences to calculate RIP indices in a method modified from the protocol originally used to study RIP patterns in *N. crassa* (Margolin *et al.*, 1998). The computer program Swaap (Pride, 2004) was used to identify observed and expected patterns of dinucleotides in a series of 200-bp sliding windows shifted in 100-bp intervals from the concatenated genomic sequence and the transposon sequences.

2.2.4 Estimation of incongruent evolutionary relationships caused by homoplasy or recombination

Since the retrotransposon *Ccret2* was found in multiple copies, with both RIPped and unmutated copies harbored in the *C. cereale* clade B genome, this element was used to determine whether RIPped sequence data could serve as reliable indicators of *C. cereale*

evolutionary relationships. PCR products were amplified using the *Ccret2*^{A15} primer pair MV-Pol-2F/MV-Pol-3R (*Ccret2*$^{POL2/3}$, Table 1) from *C. cereale* isolates PA-50005, PA-50231 and *C. sublineolum* isolate S12.001 genomic DNA. PCR products were gel purified using the Gene Clean III kit (Qbiogene, Irvine, CA), and cloned into the vector pGEM-T Easy (Promega, Madison, WI). 39 inserts from *Ccret2*$^{POL2/3}$ colonies were sequenced in both directions; the resultant nucleotide data were aligned as described above (*section 4.2.2*) and analyzed using ML in PAUP*, with the closely related, non-RIPped retrotransposon *Cg*ret from *C. gloeosporioides* (Zhu and Oudemans, 2000) used as the outgroup taxon. Starting trees were obtained from 10 random-addition replicates using a tree-bisection reconnection branch swapping algorithm. Posterior probabilities supporting the tree topology were generated using Bayesian inference as described above (*section 4.2.2*). The model of evolution that best described the dataset was estimated using Modeltest v3.06 (Posada and Crandall, 1998) and specified for both the ML and Bayesian estimates.

Several independent analyses were performed on the *Ccret2*$^{POL2/3}$ multiple sequence alignment to detect incongruence due to recombination and discriminate its influence from the effects of homoplasy induced by the RIP mutation process. First, we graphically assessed the dataset for evidence of network-like relationships between the *Ccret2*$^{POL2/3}$ copies using the split decomposition method implemented in SplitsTree v4.2 (Huson, 1998, Huson and Bryant, 2006). Because a subset of the sequences was characterized by relatively high A+T nucleotide base composition, the split decomposition analysis was conducted using LogDet distance (Lockhart *et al.*, 1994) to ensure groupings were the result of legitimate signal and not confounded by compositional biases. Next, to detect discordant relationships inferred along the length

of the *Ccret2*$^{POL2/3}$ sequences, the program SlidingBayes v0.94 (Paraskevis *et al.*, 2005), working together with MrBayes, was used to perform a Bayesian sliding window analysis of the *Ccret2*$^{POL2/3}$ alignment to identify any incongruent tree topologies. A sliding window of 50 nucleotides was run along the length of the alignment, with four MCMCMC chains run for 10^6 generations per window; a 75% consensus tree was then generated in PAUP* from the resultant 3800 trees for each of the 20 windows. The Shimodaira-Hasegawa likelihood ratio test (Shimodaira and Hasegawa, 1999), performed in PAUP* using 10,000 RELL bootstrap replicates, was employed to test whether discordant topologies in the 20 consensus trees were generated along the length of the nucleotide alignment. Third, recombination in the dataset was estimated through calculation of the Phi statistic (ϕ_w) (Bruen *et al.*, 2006) as launched through SplitsTree. Finally, possible recombination breakpoints in the multiple sequence alignment were assessed using GARD (Kosakovsky Pond *et al.*, 2006).

2.2.5 Nucleotide sequence accession numbers

All new sequences generated by this study have been deposited in the GenBank database (accession numbers **DQ663091**–**DQ663113**, **DQ663495**–**DQ663534**, **DQ666106**–**DQ666136**, **DQ666147** –**DQ666164**, **DQ667983**–**DQ668029** and **EF067890**–**EF067895**).

2.3 Results

2.3.1 Identification and nomenclature of TEs from the *C. cereale* and *C. sublineolum* genomes

A total of 35 unique transposon copies were identified using a combination of methods: (1) 5 plasmid clones were identified as containing repetitive elements due to their strong hybridization signal when probed with *C. cereale* total genomic DNA; (2) 13 TEs were found during the course of random sequencing of the PA-50005 genomic DNA library; (3) 6 TEs were sequenced from subclones of the cosmid 9F8 identified through colony hybridization; and (4) 9 *C. cereale* and 2 *C. sublineolum* retrotransposon sequences were obtained through PCR amplification and subsequent cloning. Significant levels of sequence similarity (*sections 3.2, 3.3*) strongly supported the identification of these elements as transposons. To simplify discussion in this narrative, the following names are used to describe the five new *C. cereale* TE species and one new *C. sublineolum* TE species identified in this study: (1) *Collect1* (*Colletotrichum cereale transposon 1*) is assigned to the *pogo* superfamily DNA transposon species and (2) *Collect2* (*Colletotrichum cereale transposon 2*) is used to describe the *Tc1/mariner* family TE. Retrotransposon taxa are named in accordance with the International Committee on Taxonomy of Viruses (ICTV) Code (Fauquet et al., 2004): (1) *Colletotrichum cereale Ccret1 virus (Ccret1* [*Colletotrichum cereale retrotransposon 1*]) to describe the *C. cereale Pseudoviridae* family long terminal repeat (LTR) retrotransposon; (2) *Colletotrichum cereale Ccret2 virus (Ccret2* [*Colletotrichum cereale retrotransposon 2*]) to describe the *C. cereale Metaviridae* family LTR retrotransposon; (3) *Colletotrichum cereale Ccret3 virus (Ccret3* [*Colletotrichum cereale retrotransposon 3*]) to describe the *C. cereale* non-LTR retrotransposon; and (4) *Colletotrichum sublineolum Cosret1 virus (Cosret1* [*Colletotrichum sublineolum retrotransposon 1*]) to describe the *C. sublineolum Metaviridae* family LTR retrotransposon. Individual copies of the TE species will be designated using superscript notation (e.g. *Ccret2*^{A15}).

2.3.2 Degenerate transposons are identified from the *C. cereale* clade B genome

Using a combination of BLASTX, PSI-BLAST and Conserved Domain Database (CDD) searches (Marchler-Bauer *et al.*, 2005, Marchler-Bauer and Bryant, 2004), 24 sequences from the PA-50005 clade B genome were identified as members of three distinct transposon families (Table 2): the *pogo* superfamily DNA transposon *Collect1* (Fig. 1), the *Pseudoviridae* family retrotransposon *Ccret1* (Fig. 2), and the *Metaviridae* family retrotransposon *Ccret2* (Fig. 3). Of the 24 *C. cereale* TEs, 23 were predicted to be interrupted by at least one and as many as 52 stop codons, suggesting that these elements would not function autonomously. These 23 TEs were also notable in that they displayed a marked bias for A+T nucleotides, averaging 67.62% A+T in an organism with an estimated genome A+T content of 51.03% (Fig. 4a, Table 2). Despite the degeneracy of these TEs, several of the elements had recognizable conserved domains characteristic of transposons. The retrotransposon integrase core domain (gnl|CDD|25582), characteristic of *Pseudoviridae* retrotransposons, was present in both *Ccret1*^{DBP6} and *Ccret1*^{M31}, and the conserved DDE superfamily endonuclease domain required for DNA transposition (gi|CDD|26040) was identified from the sequence of *Collect1*^{M21}.

Although some of the *C. cereale* transposons of the same species were predicted to be present in overlapping regions as inferred by positional homology, reliable contigs could not be generated from the copies due to a low level of sequence similarity. Based upon their predicted position in the *gag* region, the nucleotide sequences of *Ccret2*^{4M13}, *Ccret2*^{1J24} and *Ccret2*^{1A1} would be expected to overlap with *Ccret2*^{DBP16} (Fig. 3), but the sequences were not similar enough to suggest derivation from a single genomic locus

(>80% similarity). Similarly, despite positional homology, *Ccret1*^{M6}, *Ccret1*^{M56} and *Ccret1*^{7H14} did not possess enough sequence similarity to be drawn from a single TE copy. *Ccret1*^{M31} was 76% identical to *Ccret1*^{DBP6} over a 331-bp overlap, but only 49% of the differences between the two sequences were caused by transitions (Fig. 2). Similarly, *Collect1*^{M40} was predicted to share positional homology with *Collect1*^{I29} based on its alignment with *M. oryzae* transposases (Fig. 1), but the low level of sequence similarity (52%) showed that the two elements were not derived from the same TE copy. Thus, we predicted that although these elements are part of the same transposon species, these TEs clearly do not represent a single contiguous transposon sequence as evidenced by both differences in nucleotide base composition and the relatively low levels of sequence similarity in regions expected to overlap.

2.3.3 Intact transposon sequences are recovered from the *C. cereale* clade A genome

In contrast with the degenerate sampling of TEs recovered from the genome of the clade B isolate PA-50005, deduced translation products of the nine transposons from the *C. cereale* clade A genome (isolates PA-50231 and NJ-6340) were intact, with none interrupted by stop codons, and nucleotide compositions close to those observed in the *C. cereale* genome (Fig. 4a). Four transposon families were represented in our survey of the clade A genome: *Collect2*, a DNA transposon of the *Tc1/mariner* family, the LTR-retrotransposons *Ccret1* and *Ccret2*, and the non-LTR retrotransposon *Ccret3* (Table 2).

Because all of the transposon sequences from the *C. cereale* clade A genome were quite different from those found in the context of the clade B genome, we sought to

determine whether there were clade B-like degenerate elements in the clade A genome. PCR primers designed to amplify the region between clade B's *Ccret2*^{DBP16} and the clade A sequence *Ccret2*^{A15} (Table 1; Fig. 3) led to amplification of a 4478-bp product from the clade B isolate PA-50005 (*Ccret2*$^{Pol3/Gag11F}$), but no product in any clade A isolate, even at lower stringency annealing temperatures (data not shown), was found. Sequencing the cloned PCR amplicon revealed that *Ccret2*$^{Pol3/Gag11F}$ was characterized by numerous stop codons and deletions, similar to other TEs from the clade B genome.

In a second attempt to identify a clade B-like TE from a *C. cereale* clade A strain, the clade B *Ccret1*^{DBP6} sequence was used as a probe to screen an ordered cosmid library of clade A strain NJ-6340 (~3x coverage). Of 6,144 clones, only a single cosmid (9F8) hybridized strongly with the probe. Because clean sequence reads could not be obtained by direct sequencing the purified 9F8 cosmid DNA, the insert DNA was excised with *Not*I, then digested with *BamH*I and subcloned into plasmid pGEM3zf+. Six unique TE sequences drawn from three different species (*Ccret1*, *Ccret3* and *Collect2*) were identified on the cosmid insert. Two individual *Ccret1* sequences were also resident on the 9F8 cosmid, providing an explanation for the difficulties encountered during direct sequencing attempts. In contrast with the five *Ccret1* elements described from the clade B genome, both of the clade A-derived *Pseudoviridae* TEs possessed intact coding sequences and A+T compositions comparable to genomic levels (Fig. 4a).

Although none of the clade A TEs identified were degenerate, each of the 9 sequences was unique (Fig. 2 and 3), suggesting that each represented a unique transposon copy. In particular, although *Ccret1*$^{9F8-1787}$ and *Ccret1*$^{9F8-662}$ – sequences identified from a single cosmid clone – were predicted to overlap with each other and the other five *Ccret* copies through positional homology (Fig. 2), the sequences were too

divergent to suggest a single TE copy. The clustering of two unique *Ccret1* elements with a copy of *Collect2* and *Ccret3* suggests that the 9F8 region of the *C. cereale* genome may serve as a target site for transposition.

2.3.4 Assessment of RIP-like patterns in the transposon sequences

The observation that 21 of the 35 TEs identified in this study – all from the *C. cereale* clade B genome – were characterized by high A+T and open reading frame (ORF) interruption levels led us to hypothesize that the transposons in this lineage were being targeted and altered by the RIP mutation process (Cambareri *et al.*, 1989). A comparative RIP index analysis (Margolin *et al.*, 1998) supported this hypothesis. Overall, the genome was characterized by an A+T content of 51.03% with the ratio of all 16 possible dinucleotide combinations (observed/expected) found to span a range of 0.75 to 1.32 (Table 2; Figs. 4a, 5). In contrast, the dinucleotide sequence ratios in the high A+T clade B transposons were skewed in a manner indicative of RIP mutation. In all three of the *C. cereale* clade B transposon species, a clear pattern of increased TpA, CpT and ApG dinucleotides and decreased levels of TpG, CpA, CpG, GpA and TpC dinucleotides were detected (Table 2; Figs. 4a, 5) in sequences predicted to be RIPped based upon high A+T levels, but not in TEs with lower, genome-level A+T ratios.

The ratio of TpA to ApT dinucleotides, a signature of RIP in many fungal species (Margolin *et al.*, 1998), in the concatenated *C. cereale* genomic DNA sequence data was 0.98, while the ratio for the predicted "normal", non-RIPped transposons was 0.79 (Table 2, Fig. 4b). However, the TpA/ApT ratio for the predicted RIP-mutated clade B sequences was elevated relative to the frequencies observed in the *C. cereale* genome, with values ranging from 1.77–2.36, and 43% of these sequences characterized by ratios > 2.0.

These values are similar to those in RIPped elements in other fungal species (Fig. 4b), and supported the identification of a *C. cereale* RIP mutation process.

To further test the hypothesis of RIP mutation, we employed the RIP index formula originally used in studies of *N. crassa* (Margolin *et al.*, 1998) to compare the frequencies of the most common target of RIP-mutation in the *C. cereale* genome (CpA, CpG) to the frequencies of corresponding dinucleotide pairs of the same base composition (ApC, GpC). The ratio (CpA+CpG)/(ApC+GpC) for the *C. cereale* genome was 1.07, whereas the range for the "normal" *C. cereale* TEs averaged 1.10. In contrast, the (CpA+CpG)/(ApC+GpC) ratio for the predicted RIP-mutated transposons averaged 0.44, consistent with RIP mutated TEs from several other ascomycetes (Table 2).

Filamentous fungi in which RIP-like mutations have been characterized show a clear preference for RIP-induced C→T transitions and complementary G→A transitions, with CpA, CpG and TpG the most common target sites in the ascomycetes *N. crassa*, *P. anserina*, *F. oxysporum*, *A. fumigatus*, *A. nidulans* and *L. maculans* (Attard *et al.*, 2005, Cambareri *et al.*, 1989, Clutterbuck, 2004, Graia *et al.*, 2001, Hua-Van *et al.*, 2001). Multiple sequence alignment of both RIPped and unRIPped homologues of *Ccret2*$^{POL2/3}$ copies (*section 3.5*), revealed that cytosines in CpA and CpG and guanines in TpG are also likely targets of RIP in *C. cereale*. Of the transposon sequences predicted as RIP-mutated in this study, almost all showed decreased levels of these dinucleotides and elevated levels of the corresponding RIP-altered dinucleotide pairs (CpA→TpA; CpG→TpG→TpA; TpG→TpA) (Table 2, Fig. 5). In comparison, the pattern of bases adjacent to RIP-mutated cytosines revealed no dominant pattern of site specificity, although the frequency of nucleotides found on the 3' side of the mutated cytosine was A>>>T>C~G, while on the 5' side the pattern was T>A>C>G. Together, these data provide considerable

evidence that RIP mutation has acted upon transposons in *C. cereale* clade B; however, none of the clade A transposons identified in this study, including the 9F8 cosmid sequences that were originally detected by means of the RIPped *Ccret1*^{DBP6} probe sequence, showed any evidence of the RIP mutational process.

2.3.5 Genomic population analysis using *Ccret2*$^{POL2/3}$ sequence data

Ccret2^{A15} is the only transposon in this study that is present in multiple copies in both of the major *C. cereale* lineages (clade A and B) as determined by Southern blot analysis (data not shown). When the cloned, sequenced PCR product of *Ccret2*^{A15} ("*Ccret2*$^{POL2/3}$") was evaluated using the program Modeltest, the dataset was most accurately described by a general time reversible (GTR) model of evolution (GTR+G; A=0.3594, C=0.2328, G=0.1764, T=0.2315; A→C 4.4901, A→G 1.8908, A→T 3.6501, C→G 3.7192; α=1.1892), but partitioning of the dataset revealed that only the clade B lineage was accurately characterized by the time reversibility constraint; therefore, we used a non-GTR equivalent (HKY85+G) to avoid overparameterization. Maximum likelihood phylogenetic analysis of the 461-bp *Ccret2*$^{POL2/3}$ alignment of 39 sequences resulted in a tree topology broadly congruent with *C. cereale* phylogenetic origin; both clade A and B were recovered (Fig. 6). In a notable exception, two sequences from the genome of *C. sublineolum* isolate S12.001 (MV29 and MV21) formed a group with the *C. cereale* clade A *Ccret2*$^{POL2/3}$ elements, separate from the other 9 *C. sublineolum* sequences (posterior probability=100). Outside of this single exception, four main lineages were recovered: (1) the outgroup sequence, *Cgret*, from *C. gloeosporioides*; (2) the *C. sublineolum* *Cosret1* clade, which took the form of a polytomy; (3) the clade A-like *Ccret2*$^{POL2/3}$ lineage,

consisting of all the PA-50231 *Ccret2*$^{POL2/3}$ copies and two sequences from *C. sublineolum* in a polytomy; and (4) the clade-B-derived *Ccret2*$^{POL2/3}$ lineage, which consisted of an extremely diverse, highly mutated assemblage of sequences, all from the genome of isolate PA-50005.

Evaluation of the *Ccret2*$^{POL2/3}$ sequences for A+T content, patterns of dinucleotides, TpA to ApT ratios and skewed frequency dinucleotide (CpA+CpG)/(ApC+GpC) ratios identified 14 of the sequences as RIP-mutated (Table 2), despite the fact that these sequences were generated using PCR primers designed from a non-RIPped sequence (*Ccret2*^{A15}). All 14 of the RIPped sequences were acquired from the genome of clade B isolate PA-50005; only a single copy from this isolate (MV25) was non-RIPped (Table 2) and placed as the basal lineage in the otherwise RIP-mutated clade B lineage. All transposon copies sequenced from the PA-50231 (clade A) and *C. sublineolum* S12.001 genomes were "normal"; i.e., there was no evidence of RIP-like alterations to the sequences.

The *Ccret2*$^{POL2/3}$ dataset was evaluated for the presence of conflicting signal indicative of either homoplasy or recombination using four individual methods. Each of the analyses identified incongruence from the *C. cereale* clade B lineage, but the clade A and *C. sublineolum*-derived sequences were accurately depicted in a bifurcating tree topology consistent with the absence of both recombination and homoplasy. First, the split decomposition network visualized using LogDet distance was tree-like except among the RIPped clade B copies, where multiple incompatibilities were detected (Fig. 7); this analysis also confirmed that the clustering of taxa was the result of phylogenetic signal rather than nucleotide compositional bias. The second test, calculation of the Phi statistic (ϕ_w), confirmed the partitioned presence of recombination in the dataset and identified

statistically significant levels of recombination in clade B both with and without the non-RIPped MV25 (P=0.02 and 1.87 E-4, respectively), but in all other clades, either individually or combined, and the dataset as a whole, the signature of recombination was not observed (P=1.0). The third analysis for incongruent signal, a Bayesian analysis of 20 datasets constructed using a sliding window approach, recovered identical branching patterns among the *C. sublineolum* and *C. cereale* clade A elements across all 20 trees, while the clade B lineage was inconsistently structured from tree to tree (not shown). The likelihoods of the 20 trees, when used to perform an SH- likelihood ratio test, identified statistically significant levels of discordance between the topologies, indicative of either recombination or homoplasy (Table 3). Likewise, when the dataset was analyzed using the GARD multiple breakpoint method, a single recombination breakpoint was detected from the clade B sequences, while the remainder of the dataset presented no evidence for recombination.

2.4 Discussion

From an organismal standpoint, the lineage-specific distribution of RIP mutation in *C. cereale* – absent in clade A, present in clade B – is an important contribution to our understanding of how this species has evolved. The widespread identification of RIPping in diverse ascomycete species, including the closely related *C. falcatum* (J.A. Crouch and B.I. Hillman; unpublished data*)*, strongly discourages the conclusion that the clade A genome might be "RIP-free", but our inability to detect a RIP signature clearly emphasizes the distinct nature of *C. cereale*'s main lineages and is consistent with previous observations about these groups. Using multilocus haplotype networks, Crouch *et al.*

(2006) determined that *C. cereale* clade A is likely derived from a single ancestral haplotype. Such a population bottleneck, through the elimination of RIP-mutated alleles, could easily have produced the pattern of absentee RIP mutated TEs observed in clade A, particularly if these sequences were of no selective advantage to the fungus. Furthermore, the many copies of RIP mutated TEs in the clade B genome is entirely consistent with the reticulate population structure inferred for clade B using haplotype networks (Crouch *et al.* 2006), since both datasets suggest the same conclusion: that clade B has been influenced by gene flow processes. Gene flow as inferred through the haplotype network alone could have resulted from either sexual recombination or parasexually-derived mitotic recombination, but the detection of RIP-mutation in the clade B sequences – a process that occurs only during meiosis – provides compelling evidence of sexual recombination in the clade B lineage, where 14 of the 15 unique transposon sequences were RIPped. Moreover, the identification of at least five diverse *Ccret2*$^{\text{POL2/3}}$ groups structured along a distinct cline of RIP-mutation levels (Fig. 6) establishes that sexual recombination was not restricted to a single isolated event; instead, the sexual morph of the fungus must have been produced on multiple occasions, at least in the clade B lineage. But an important question leads to the next logical area of inquiry for this species: did the inferred clade B sexual recombination occur cryptically in extant populations or is the observed RIP signature merely the artifact of a long defunct sexual ancestral state?

For the genus *Colletotrichum* and its teleomorph *Glomerella*, the genetic basis of sexual compatibility is uncertain, despite the pioneering experiments of Edgerton, Wheeler and colleagues between 1914-1959 that served to elucidate the genetics of sexual development in *G. cingulata* (e.g. Driver and Wheeler, 1955; Edgerton, 1914; Edgerton *et*

al., 1945; Lucas *et al.*, 1944; Wheeler *et al.*, 1948; Wheeler *et al.*, 1959). What is well known, however, is that the regulation of mating in *Colletotrichum* is often unpredictable and appears to be rather different from that in other filamentous ascomycetes. While sexual recognition for most of the ascomycota is controlled by a one locus, two allele (idiomorph) mating system, generation of the *Glomerella* state is thought to be governed by at least two unlinked loci (e.g. *G. graminicola;* Vaillancourt *et al.*, 2000) or by means of a single locus with multiple alleles (e.g. *G. cingulata*; Cisar and TeBeest, 1999). To date, extensive sampling of 11 *Colletotrichum* species, including *C. cereale*, has resulted in the detection of only a single idiomorph – the *MAT1-2*, regardless of whether the species in question is heterothallic or homothallic (Crouch *et al.*, 2006; Chen *et al.*, 2002; Du *et al.*, 2005; Rodriguez-Guerra *et al.*, 2005; J.A. Crouch and B.I. Hillman, unpublished data). Even the prediction of appropriate mating partners for genetic analysis in the genus can be difficult, since many species, including *G. graminicola*, *G. cingulata* and *G. lindemuthiana*, are both heterothallic and homothallic (Cisar *et al.*, 1994; Cisar and TeBeest, 1999; Rodriguez-Guerra *et al.*, 2005; Wheeler, 1954). Furthermore, individual isolates are frequently infertile, irrespective of mating conditions. Thus for *Colletotrichum*, the indirect assessment of recombination by means of molecular analysis, rather than through the direct experimental or natural observation of biological mating, is likely to yield a more accurate assessment of whether recombination has influenced populations of the fungus. The observation of RIP mutation in the present study suggests a greater level of complexity in populations of *C. cereale* than previously suspected, with the available evidence favoring the view that the evolution of *C. cereale* has been driven, at least in part, by recombination.

From a broad evolutionary perspective, the amino acid sequences of the transposons identified in this study – even those that are RIP mutated – reflect phylogenetic relationships that are generally consistent with currently accepted taxonomy (Figs. 1, 2, 3). But as the analysis moved to the nucleotide level to examine the *Ccret2*$^{POL2/3}$ populations within individual genomes, inconsistencies began to emerge. The first irregularity, in which two of the 11 copies of *Ccret2*$^{POL2/3}$ sequenced from *C. sublineolum* were identical to copies of *Ccret2* from *C. cereale* clade A, might have resulted due to a retained ancestral polymorphism since it has been determined that *Ccret2* was present in the common ancestor of the sister species *C. cereale*, *C. sublineolum* and *C. falcatum* (J.A. Crouch, B.I. Hillman; unpublished data). Different copies of the TE might then have been randomly distributed during the subsequent speciation process. Horizontal gene transfer, which has been systematically established in only a few instances for eukaryotic TEs (Jordan *et al.*, 1999, Daniels *et al.*, 1990, Diao *et al.*, 2006), provides an alternative explanation, and has been proposed to account for the unexpected presence of an active *Tad* element in *N. crassa* (Anderson *et al.*, 2001).

A second irregularity appeared in the phylogenetic analysis of the *Ccret2*$^{POL2/3}$ dataset: the RIPped clade B *Ccret2*$^{POL2/3}$ elements clustered into 5 main groups, with each group presumably representing different numbers of RIP-mutation cycles acting on different copies of the element. The branches constructed from the RIPped TE copies defy cladistic assumptions, since these elements are established in the reverse chronological order of the actual evolutionary path within the tree topology. The most divergent, and in this case, also the oldest copies of *Ccret2*$^{POL2/3}$ are found at the tips of the tree, since the oldest copies of the element would have undergone repeated rounds of mutation relative to more recently inserted, less RIPped copies of the transposon (Fig. 6).

In this scenario, although the basal taxa in the *Ccret2*$^{POL2/3}$ tree possess the more ancestral-like sequence, these transposon copies are actually the youngest examples of RIP-mutated elements within the sample.

The third anomaly in the *Ccret2*$^{POL2/3}$ dataset was concentrated within the RIPped clade B group, where numerous inconsistencies were observed. Several incongruent clade B topologies were recovered when the dataset was subjected to sliding window analysis, suggesting either the presence of recombination or RIP-induced homoplasy. Recombination was independently supported by the ϕ_w analysis, but since only a single recombination breakpoint was detected using the GARD method, it appears that the numerous incongruent topologies recovered using the sliding window analysis were the result of RIP mutation-derived substitution and rate heterogeneity rather than high levels of recombination. Our data cannot rule out the potential for recombination among the RIPped *Ccret2* elements, since recombination can be a powerful force acting on retrotransposons, as exemplified in *Saccharomyces* (for review, see Mieczkowski *et al.*, 2006). The hypermutation of the RIPped sequences might even have diluted the signal of recombination in this dataset. Conversely, the accumulation of TEs in non-recombining regions of the genome is predicted both by population genetics theory (Montgomery *et al.*, 1991, Langley *et al.*, 1988, Charlesworth *et al.*, 1994) and is also a common trend in many diverse organisms, including *Drosophila melanogaster*, *Tetraodon nigrovidis* and *Arabidopsis thaliana* (Fischer *et al.*, 2004, Langley *et al.*, 1988, Maside *et al.*, 2005, Kapitonov and Jurka, 2003, Wright *et al.*, 2003). In our dataset, it appears that RIP mutation-induced homoplasy rather than recombination is the primary driver of conflict.

In this study, the data clearly showed that the accelerated mutations experienced during the RIP process are not equivalent to the sequence variation experienced by their

"normal" non-RIPped counterparts, with the rate of evolution experienced by different RIPped copies within a single *C. cereale* strain greatly exceeding the mutational forces experienced, even when compared to that which occurred on the intraspecific level between *C. cereale* and *C. sublineolum*. Phylogenetic analysis of the *Ccret2*$^{POL2/3}$ copies visibly illustrated the high levels of divergence of the RIPped TEs relative to non-mutated elements of the same mobile element (Fig. 6); the normal TE sequences showed little variability, but for the RIPped group, the tree topology was characterized by exceptionally long branches. Even if RIP is no longer an active defense in the genome of *C. cereale* clade B, once these elements are RIP-altered and inactivated, the selective pressures experienced by these genomic relics are likely to be no longer equivalent to those encountered by their non-mutated counterparts and the RIPped elements would evolve under an entirely different set of circumstances than their functional counterparts.

Comprehensive genome-scale studies of fungal transposon ecology and evolution have been largely confined to the yeasts, despite the demonstrated power of TEs to advance our understanding of the biology and evolution across a wide range of eukaryotes. Noteworthy discoveries that have greatly expanded our knowledge of genome variability and evolution have resulted from studies of TEs in plants, insects, mammals, and, in particular, humans (e.g. Brouha *et al.*, 2003, Du *et al.*, 2006, Nikaido *et al.*, 2001, Roy-Engel *et al.*, 2002, Vieira and Biemont, 2004, Vincent *et al.*, 2003). RIP mutation will present a complicating factor in many fungal TE evolutionary analyses, as established both in the current study and in studies of *F. oxysporum* and *M. violaceum* (Hua-Van *et al.*, 2001, Hood *et al.*, 2005), but this is more than offset by the wealth of information gained. With the increasing availability of fungal genome sequence data (97 fungal genome sequencing projects are currently curated by the NCBI Entrez Genome

Project), we have unprecedented opportunities to explore and interpret transposon distribution, function and diversity.

2.5 References

Abascal, F., R, Z. & Posada, D. (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* **21:** 2104-2105.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215:** 403-410.

Anderson, C., Tand, Q. & Kinsey, J. (2001) Elimination of active *Tad* elements during the sexual phase of *Neurospora crassa* life cycle. *Fungal Genet. Biol.* **33:** 49-57.

Attard, A., Gout, L., Ross, S., Parlange, S., Cattolico, L., Balesdent, M.-H. & Rouxel, T. (2005) Truncated and RIP-degenerated copies of the LTR retrotransposon *Pholy* are clustered in a pericentromeric region of the *Leptosphaeria maculans* genome. *Fungal Genet. Biol.* **42:** 30-41.

Bender, J. & Kleckner, N. (1992) Tn10 insertion specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequences. *Proc. Natl. Acad. Sci. USA* **89:** 7996-8000.

Bhat, A., Tamuli, R. & Kasbekar, D. (2004) Genetic transformation of *Neurospora tetrasperma*, demonstration of repeat-induced point mutation (RIP) in self-crosses and a screen for recessive RIP-defective mutants. *Genetics* **167:** 1155-1164.

Bouvet, G.F., Jacobi, V. & Bernier, L. (2007) Characterization of three DNA transposons in the Dutch elm disease fungi and evidence of repeat-induced point (RIP) mutations. *Fungal Genet. Biol.* **44**: 430-443.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. & Kazazian Jr., H.H. (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. USA* **100:** 5280-5285.

Bruen, T.C., Phillipe, H. & Bryant, D. (2006) A simple and robust statistical test to detect the presence of recombination. *Genetics* **172:** 2665-2681.

Cambareri, E.B., Jensen, B.C., Schabtach, E. & Selker, E.U. (1989) Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244:** 1571-1575.

Cambareri, E.B., Singer, M.J. & Selker, E.U. (1991) Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. *Genetics* **127:** 699-710.

Chalker, D.L. & Sandmeyer, S.B. (1992) Ty3 integrates within the region of RNA polymerase III transcription initiation. *Genes Dev.* **6:** 117-128.

Chalmers, R., Guhathakurta, A., Benjamin, H. & Kleckner, N. (1998) IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. *Cell* **93:** 897-908.

Charlesworth, B., Sniegowski, P. & Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* **371:** 215-220.

Chen, F., Goodwin, P.H., Khan, A., Hsiang, T. 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris*. *Can. J. Microbiol.* **48:** 427-436.

Cisar, C.R., Spiegel, F.W., TeBeest, D.O., Trout, C. 1994. Evidence for mating between isolates of *Colletotrichum gloeosporioides* with different host specificities. *Curr. Genet.* **25:** 330-335.

Cisar, C.R., TeBeest, D.O. 1999. Mating system of the filamentous ascomycete, *Glomerella cingulata*. *Curr. Genet.* 35**:** 127-133.

Clutterbuck, A.J. (2004) MATE transposable elements in *Aspergillus nidulans*: evidence of repeat-induced point mutation. *Fungal Genet. Biol.* **41:** 308-316.

Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U. & Macino, G. (1996) Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA-DNA methylation. *EMBO J.* **15:** 3153-3163.

Crouch, J.A., Clarke, B.B. & Hillman, B.I. (2005) Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. *Int. Turfgrass Soc. Res. J.* **10:** 186-195.

Crouch, J.A., Clarke, B.B. & Hillman, B.I. (2006) Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. *Phytopathology* **96:** 46-60.

Daniels, S., Peterson, K., Strausbaugh, L., Kidwell, M. & Chovnick, A. (1990) Evidence for horizontal transmission of the P transposable element between *Drosophila* species. *Genetics* **124:** 339-355.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N.D., Lee, Y.-H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E. & Birren, B.W. (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* **434:** 980-986.

Devine, S.E. & Boeke, J.D. (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes Dev.* **10:** 620-633.

Diao, X., Freeling, M. & Lisch, D. (2006) Horizontal transfer of a plant retrotransposon. *PLoS Biology* **4:** 0119-0128.

Driver, C., Wheeler, H. 1955. A sexual hormone in *Glomerella*. *Mycologia* **47,** 311-316.

Du, M., Schardl, C. L., Nuckles, E.M. & Vaillancourt, L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of Colletotrichum species complexes. *Mycologia* **97**: 641-658.

Du, C., Swigonova, Z. & Messing, J. (2006) Retrotranspositions in orthologous regions of closely related grass species. *BMC Evol. Biol.* **6:** 62.

Edgerton, C. 1914. Plus and minus strains in the genus *Glomerella*. *Am. J. Bot.* **1,** 244-254.

Edgerton, C., Chilton, S., Lucas, G. 1945. Genetics of *Glomerella*. II. Fertilization between strains. *Am. J. Bot.* 32**:** 115-118.

Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U., and Ball, L. A. (2004) *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses,* Elsevier/Academic Press, New York.

Felsenstein, J. (2006) PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Fischer, C., Bouneau, L., Coutanceau, J., Weissenbach, J., Volff, J. & Ozouf-Costaz, C.

(2004) Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish *Tetraodon nigroviridis*. *Gene* **336:** 175-183.

Galagan, J.E. & Selker, E.U. (2004) RIP: the evolutionary cost of genome defense. *Trends in Genetics* **20:** 417-422.

Gardner, M., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R., Carlton, J., Pain, A., Nelson, K., Bowman, S., Paulsen, I., James, K., Eisen, J., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M., Vaidya, A., Martin, D., Fairlamb, A., Fraunholz, M., Roos, D., Ralph, S., McFadden, G., Cummings, L., Subramanian, G., Mungall, C., Venter, J., Carucci, D., Hoffman, S., Newbold, C., Davis, R., Fraser, C. & Barrell, B. (2002) Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419:** 498-511.

Goffeau, A., Barrell, B., Bussey, H., Davis, R., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J., Jacq, C., Johnston, M., Louis, E., Mewes, H., Murakami, Y., Philippsen, P., Tettelin, H. & Oliver, S. (1996) Life with 6000 genes. *Science* **274:** 563-567.

Graia, L., Lespinet, O., Rimbault, B., Dequard-Chablet, M., Choppin, E. & Picard, M. (2001) Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. *Mol. Microbiol.* **40:** 586-595.

Hood, M.E., Katawczik, M. & Giraud, T. (2005) Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. *Genetics* **170:** 1081-1089.

Hua-Van, A., Langin, T. & Daboussi, M.-J. (2001) Evolutionary history of the *impala* transposon in *Fusarium oxysporum*. *Mol. Biol. Evol.* **18:** 1959-1969.

Hua-Van, A., Le Rouzic, A., Maisonhaute, C. & Capy, P. (2005) Abundance, distribution and dynamics of retrotransposons and transposons: similarities and differences. *Cytogenet. Genome Res.* **110:** 426-440.

Huelsenbeck, J.P. & Ronquist, F. (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19:** 1572-1574.

Huson, D.H. (1998) SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14:** 68-73.

Huson, D.H. & Bryant, M.L. (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23:** 254-267.

IHGSC (2001) Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature* **409:** 860-921.

Ikeda, K.-I., Nakayashiki, H., Kataoka, T., Tamba, H., Hashimoto, Y., Tosa, Y. & Mayama, S. (2002) Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. *Mol. Microbiol.* **45:** 1355-1364.

IMGSC (2001) Initial sequencing and comparative analysis of the mouse genome. International Mouse Genome Sequencing Consortium. *Nature* **420:** 520-562.

Jordan, I.K., Matyunina, L.V. & McDonald, J.F. (1999) Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. *Proc. Natl. Acad. Sci. USA* **96:** 12621-12625

Kapitonov, V.V. & Jurka, J. (2003) Molecular paleontology of transposable elements in

the *Drosophila melanogaster* genome. *Proc. Natl. Acad. Sci. U.S.A.,* **100,** 6569-6574..

Ketting, R., Fisher, S. & Plasterk, R. (1997) Target choice determinants of the Tc1 transposon of *Caenorhabditis elegans*. *Nucleic Acids Res.* **25:** 4041-4047.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. & Frost, S.D.W. (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23:** 1891-1901.

Langley, C., Montgomery, E., Hudson, R.R., Kaplan, N. & Charlesworth, B. (1988) On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res.* **52:** 223-235.

Liu, G., Geurts, A., Yae, K., Srinivasan, A., Fahrenkrug, S., Largaespada, D., Takeda, J., Horie, K., Olson, W. & Hackett, P. (2005) Target-site preferences of *Sleeping Beauty* transposons. *J. Mol. Biol.* **346:** 161-173.

Lockhart, P.J., Steel, M.A., Hendy, M.D. & Penny, D. (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. *Mol. Biol. Evol.* **11:** 605-612.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J. & Bryant, S.H. (2005) CDD: a Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33:** D192-196.

Lucas, G., Chilton, S., Edgerton, C. 1944. Genetics of *Glomerella*. I. Studies on the behavior of certain strains. *Am. J. Bot.* **31:** 233-239.

Marchler-Bauer, A. & Bryant, S.H. (2004) CD-Search: protein domain annotations on the fly. *Nucleic Acids Res.* **32:** W327-331.

Margolin, B.S., Garrett-Engele, P.W., Stevens, J.N., Fritz, D.Y., Garrett-Engele, C.,

Metzenberg, R.L. & Selker, E.U. (1998) A methylated *Neurospora* 5S pseudogene contains a transposable element inactivated by repeat-induced point mutation. *Genetics* **149:** 1787-1797.

Maside, X., Assimacopoulos, S. & Charlesworth, B. (2005) Fixation of transposable elements in the *Drosophila melanogaster* genome. *Genet. Res.* **85:** 195-203.

Messing, J. & Dooner, H. (2006) Organization and variability of the maize genome. *Curr. Opin. Plant Biol.* **9:** 157-163.

Mieczkowski, P., Lemoine, F. & Petes, T. (2006) Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. *DNA Repair* **5:** 1010-1020.

Montgomery, E., Huang, S., Langley, C. & Judd, B. (1991) Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics,* **129,** 1085-1098.

Montiel, M., Lee, H. & Archer, D. (2006) Evidence of RIP (repeat-induced point mutation) in transposase sequences of *Aspergillus oryzae*. *Fungal Genet. Biol.* **43:** 439-445.

Nakayashiki, H., Nishimoto, N., Ikeda, K., Tosa, Y. & Mayama, S. (1999) Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. *Mol. Gen. Genet.* **261:** 958-966.

Neuveglise, C., Sarfati, J., Latge, J.P. & Paris, S. (1996) Afut1, a retrotransposon-like

element from Aspergillus fumigatus. *Nucleic Acids Res.,* **24,** 1428-1434.

Nielson, M., Hermansen, T. & Aleksenko, A. (2001) A family of DNA repeats in *Aspergillus nidulans* has assimilated degenerated retrotransposons. *Mol. Gen. Genet.* **265:** 883-887.

Nikaido, M., Matsuno, F., Hamilton, H., Brownell, R.J., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A., Fordyce, R., Hasegawa, M. & Okada, N. (2001) Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* **98:** 7384-7389.

Paraskevis, D., Deforche, K., Lemey, P., Magiorkinis, G., Hatzakis, A. & Vandamme, A.-M. (2005) SlidingBayes: exploring recombination using a sliding window approach based on Bayesian phylogenetic inference. *Bioinformatics* **21:** 1274-1275.

Posada, D. & Crandall, K.A. (1998) Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14:** 817-818.

Pride, D. (2004) Swaap 1.0.1: a tool for analyzing substitutions and similarity in multiple alignments. Distributed by the author.

Rodriguez-Guerra, R., Ramirez-Rueda, M., Cabral-Enciso, M., Garcia-Serrano, M., Lira-Maldonado, Z., Guevara-Gonzalez, R., Gonzalez-Chavira, M., Simpson, J. 2005. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. *Mycologia* **97:** 793-803.

Roy-Engel, A.M., Carroll, M.L., El-Sawy, M., Salem, A.-H., Garber, R.K., Nguyen, S.V., Deininger, P.L. & Batzer, M.A. (2002) Non-traditional Alu evolution and primate genomic diversity. *J. Mol. Biol.* **316:** 1033-1040.

Sambrook, J. & Russell, D. (2001) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York.

Selker, E.U., Cambareri, E.B., Jensen, B.C. & Haack, K.R. (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell* **51:** 741-752.

Shimodaira, H. & Hasegawa, M. (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16:** 1114-1116.

Shiu, P.K., Raju, N.B., Zickler, D. & Metzenberg, R.L. (2001) Meiotic silencing by unpaired DNA. *Cell* **107:** 905-916.

Singleton, T.L. & Levin, H.L. (2002) A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. *Eukaryot. Cell,* **1,** 44-55.

Smiley, R.W., Dernoeden, P.H. & Clarke, B.B. (2005) *Compendium of Turfgrass Diseases*, APS Press, St. Paul.

Swofford, D.L. (2000) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. *Nucleic Acids Res.,* **22,** 4673-4680.

Thon, M., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T. & Dean, R. (2006) The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol.* **7:** R16.

Vieira, C. & Biemont, C. (2004) Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. *Genetica* **120:** 115-123.

Vincent, B.J., Myers, J.S., Ho, H.J., Kilroy, G.E., Walker, J.A., Watkins, W.S., Jorde,

L.B. & Batzer, M.A. (2003) Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. *Mol. Biol. Evol.* **20:** 1338-1348.

Watters, M.K., Randall, T.A., Margolin, B.S., Selker, E.U. & Stadler, D.R. (1999) Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. *Genetics* **153:** 705-714.

Wheeler, H. 1954. Genetics and evolution of heterothallism in *Glomerella*. *Phytopathology* 44**:** 342-345.

Wheeler, H., Driver, C., Campa, C. 1959. Cross- and self-fertilization in *Glomerella*. *Am. J. Bot.* **46:** 361-365.

Wheeler, H., Olive, L., Ernest, C., Edgerton, C. 1948. Genetics of *Glomerella*: V. Crozier and ascus development. *Am. J. Bot.* **35:** 722-728.

Wicker, T., Robertson, J., Schulze, S., Feltus, F., Magrini, V., Morrison, J., Mardis, E.,
Wilson, R., Peterson, D., Paterson, A. & Ivarie, R. (2005) The repetitive landscape of the chicken genome. *Genome Res.* **15:** 126-136.

Wright, S., Gagrawal, N. & Bureau, T. (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res.* **13:** 1897-1903.

Zhu, P. & Oudemans, P.V. (2000) A long terminal repeat retrotransposon *Cg*ret from the phytopathogenic fungus *Colletotrichum gloeosporioides* on cranberry. *Curr. Genet.* **38:** 241-247.

Zou, S., Ke, N., Kim, J.M. & Voytas, D.F. (1996) The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. *Genes Dev*. **10:** 634-645.

Table 2.1

Primer sequences used in this study.

Gene	Region	Library (lineage)	Lineages amplified	Primer Name	Sequence (5'→ 3')
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-POL-1R	5' CGTACGGTCCATGCTCTG 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-POL-2F	5' CAGAGCATGGACCGTACG 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-POL-3R	5' CAGTACCTTGTGTATGTG 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A	MV-GAG-1R	5' TCTGCATTCGTCGTAGAG 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A	MV-GAG-2F	5' CTCTACGACGAATGCAGA 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A	MV-GAG-3F	5' CGAGCAAAATCGAACGAA 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-RET-1F	5' AAGGCTGCATTACACTACG 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-RET-2R	5' CAGGCGTGGAGTTCTTT 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-GAG-5F	5' AATCCTTAGTCTTTATGTTCT 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-GAG-6R	5' TTATTATTACGCTAGTTATTATTT 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-7F	5' CACTACGAAGGCAAAGCACAC 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-8R	5' TATCACATCCAAGCGTCCTATCT 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-11F	5' GTGCTATAACTGTAAGAAGAT 3'
C-cret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-12R	5' CAGTTGGCGTTTGTCGTT 3'
C-cret2 LTR retrotransposon	DBP16	PA-50005 (B)	B	MV-GAG-20F	5' GCTTAGTAGTAAAGTTAAG 3'
C-cret2 LTR retrotransposon	DBP16	PA-50005 (B)	B	MV-GAG-21R	5' CTGTAATGTTAAGTCTAG 3'
C-cret1 LTR retrotransposon	DBP6	PA-50005 (B)	B	PV-INT-40F	5' AGGGCTGTGTCAATACTCA 3'
C-cret1 LTR retrotransposon	DBP6	PA-50005 (B)	B	PV-INT-41R	5' GTCTTCCCTTCCACTGTTA 3'
Collect1 DNA transposon	I-29	PA-50005 (B)	B	*pogo* -20F	5' GGTAGGTATGCCTTATAC 3'
Collect1 DNA transposon	I-29	PA-50005 (B)	B	*pogo* -21R	5' CCTTCTAATACTTACTTAG 3'

Table 2.2

Dinucleotide frequencies.

Transgene	RIP?	Isolate (clade)	Length (bp)	A+T%	RIP indices			Dinucleotide base frequencies (observed/expected)															Accession
								AA	TT	AC	TG	AG	TC	AT	TA	CA	GT	CC	GG	TG	GC	CT	GA	
Gal4vp7 M48	–	PA-90005 (B)	699	42.91	0.58	0.59	0.46	0.71	0.55	1.26	0.57	0.62	1.76	1.61	0.59	0.83	1.12	0.73	0.73	1.50	0.70	1.03	1.38	DQ661507
Gal4vp7 898-1485	–	NA4140 (A)	1,485	46.65	0.87	1.17	1.03	1.06	1.36	0.91	0.99	0.98	0.94	0.91	0.79	0.85	0.62	0.93	0.89	1.06	1.09	1.12	1.16	HM367895
Gal4vp7 898-1256	–	NA4140 (A)	1,256	49.88	0.83	1.69	3.02	0.74	0.68	0.62	1.11	1.20	0.89	1.38	1.14	1.16	0.72	1.39	1.14	0.41	1.24	1.13	0.86	HM367891
Gal4 898-1767	–	NA4140 (A)	1,767	43.80	0.65	1.16	0.92	1.03	1.23	0.84	1.32	1.32	1.05	1.07	0.70	0.95	0.84	1.09	0.86	1.06	0.97	0.82	1.17	HM367893
Gal4 898-462	–	NA4140 (A)	462	43.35	0.76	1.13	0.65	1.05	1.05	0.89	0.94	1.00	1.04	1.09	0.83	0.83	0.68	0.94	0.88	1.17	0.76	1.01	1.21	HM367899
Gal2 M22	–	PA-90005 (B)	641	50.13	0.93	0.96	0.77	0.88	1.38	1.10	0.79	0.98	0.99	0.84	0.78	0.89	0.64	0.91	1.01	1.07	0.82	1.09	1.24	DQ661503
Gal2 M5	–	PA-90231 (A)	1,000	50.03	0.90	0.99	0.78	0.87	1.14	1.08	0.76	0.96	1.15	0.89	0.80	0.94	0.64	0.92	0.92	1.06	0.83	1.15	1.27	DQ661513
Gal2 POL23 M12	–	PA-90231 (A)	461	51.38	0.89	0.96	0.76	0.86	1.13	1.07	0.79	0.99	1.22	0.88	0.78	0.94	0.72	0.94	0.96	1.10	0.84	1.15	1.26	DQ661517
Gal2 POL23 M911	–	PA-90231 (A)	461	50.37	0.81	1.02	0.75	0.89	1.12	1.06	0.80	0.94	1.22	0.91	0.74	0.95	0.64	0.91	0.92	1.08	0.81	1.13	1.28	DQ661518
Gal2 POL23 M928	–	PA-90005 (B)	461	50.13	0.93	0.96	0.77	0.88	1.38	1.10	0.79	0.98	0.99	0.84	0.78	0.89	0.64	0.91	1.01	1.07	0.82	1.09	1.24	DQ661510
Gal4 898-2137	–	NA4140 (A)	2,137	46.37	0.80	1.16	1.08	1.05	1.11	0.88	0.89	1.10	1.23	0.89	0.71	0.98	0.73	1.16	0.73	0.93	1.03	1.18	1.09	DQ667894
Gal4 898-1518	–	NA4140 (A)	1,518	48.26	0.57	1.18	1.02	0.85	1.11	1.01	1.16	1.02	1.12	0.98	0.76	1.17	0.96	0.91	0.82	0.84	0.86	0.92	1.20	HM367892
Gam7 POL23 M921	–	S12001 (C, subcluster)	461	51.37	0.91	0.98	0.74	0.93	0.46	1.08	0.81	0.97	1.15	0.89	0.81	0.93	0.73	0.88	0.93	1.08	0.79	1.14	1.31	DQ661506
Gam7 POL23 M929	–	S12001 (C, subcluster)	461	52.95	0.83	1.46	0.99	0.93	1.25	1.13	1.21	0.97	1.04	0.91	0.57	0.99	0.41	0.81	0.91	0.86	0.85	1.40	1.43	DQ661513
Gal4vp7 129	RIP	PA-90005 (B)	587	70.98	2.11	0.27	1.47	0.84	0.92	1.27	0.23	1.66	0.58	0.65	1.37	0.37	1.00	1.28	0.90	0.46	0.68	1.81	1.07	DQ661509
Gal4vp7 VIII	RIP	PA-90005 (B)	600	66.33	2.11	0.36	17.38	0.65	0.87	1.24	0.47	1.71	0.43	0.83	1.69	0.39	1.13	1.49	1.02	0.01	0.91	1.84	0.78	DQ661502
Gal1 2B96	RIP	PA-90005 (B)	2,889	76.32	1.84	0.08	0.91	0.78	0.75	1.31	0.06	1.78	0.45	0.83	1.46	0.11	0.85	1.28	0.98	0.52	0.47	2.19	1.06	DQ661510
Gal1 TH14	RIP	PA-90005 (B)	641	74.31	2.01	0.19	3.47	0.70	0.82	1.45	0.80	2.12	0.48	0.75	1.50	0.50	1.23	1.10	0.42	0.15	0.51	1.75	0.93	DQ661499
Gal1 M6	RIP	PA-90005 (B)	624	68.19	2.21	0.32	4.75	0.76	0.83	0.97	0.25	1.82	0.76	0.68	1.53	0.38	0.98	1.20	1.22	0.25	1.21	1.83	0.85	DQ661501
Gal1 M31	RIP	PA-90005 (B)	582	57.94	1.86	0.75	0.74	0.50	1.16	1.01	0.47	1.49	0.83	0.67	1.25	0.69	0.54	0.44	0.85	1.56	1.16	1.56	1.11	DQ661505
Gal1 M56	RIP	PA-90005 (B)	418	73.94	1.80	0.23	1.96	0.89	0.79	0.83	0.36	1.76	1.11	0.79	1.42	0.13	1.33	1.08	1.21	0.41	0.79	1.88	0.41	DQ661508
Gal1 BB716	RIP	PA-90005 (B)	2,380	67.01	2.19	0.55	1.82	0.78	1.17	1.39	0.45	1.52	0.35	0.59	1.29	0.68	0.67	0.80	0.89	0.71	1.30	1.89	1.06	DQ661511
Gal2 1A1	RIP	PA-90005 (B)	586	57.09	1.44	0.60	1.80	0.80	1.88	1.24	0.33	1.52	0.48	0.65	1.25	0.80	0.68	0.87	0.92	0.68	1.22	1.51	0.99	DQ661495
Gal2 4M13	RIP	PA-90005 (B)	335	66.67	1.92	0.56	2.35	0.80	1.19	1.22	0.26	1.62	0.49	0.67	1.29	0.81	0.68	0.92	0.98	0.53	1.39	1.69	0.89	DQ661498
Gal2 M3	RIP	PA-90005 (B)	689	69.62	1.80	0.39	1.74	0.84	1.03	1.21	0.13	1.58	0.48	0.72	1.36	0.64	0.73	1.24	1.46	0.54	0.84	1.83	0.97	DQ661500
Gal2 4M16	RIP	PA-90005 (B)	944	61.34	1.77	0.77	1.73	0.61	1.26	1.19	0.49	1.44	0.70	0.61	1.68	0.95	0.69	0.79	0.94	0.64	1.11	1.43	1.04	DQ661497
Gal2 M26	RIP	PA-90005 (B)	192	70.41	2.04	0.25	1.73	0.75	0.73	1.21	0.45	1.71	0.92	0.78	1.39	0.18	1.24	0.73	1.71	0.39	0.67	1.71	0.31	DQ661506
Gal2 1M4	RIP	PA-90005 (B)	619	71.34	2.26	0.20	8.73	0.83	0.86	0.75	0.24	2.23	0.99	0.66	1.50	0.12	1.13	1.09	0.93	0.16	1.44	1.87	0.50	DQ661496
Gal5 PolIIGag11	RIP	PA-90005 (B)	4,478	63.97	1.93	0.58	2.89	0.87	0.99	1.15	0.54	1.53	0.61	0.70	1.35	0.65	0.91	1.06	1.13	0.38	1.10	1.67	0.85	DQ661512
Gal2 POL23 M934	RIP	PA-90005 (B)	461	64.06	1.87	0.52	2.35	0.77	1.16	1.20	0.19	1.69	0.43	0.73	1.39	0.80	0.69	0.93	0.83	0.57	1.35	1.69	0.97	DQ661523
Gal2 POL23 M940	RIP	PA-90005 (B)	461	66.67	1.92	0.56	2.35	0.80	1.19	1.22	0.26	1.62	0.49	0.67	1.29	0.81	0.68	0.92	0.98	0.53	1.39	1.69	0.89	DQ661526
Gal2 POL23 M947	RIP	PA-90005 (B)	461	66.62	1.90	0.59	1.74	0.80	1.03	1.25	0.13	1.58	0.48	0.72	1.36	0.64	0.73	1.04	1.46	0.54	0.84	1.63	0.97	DQ661528
Gal2 POL23 M958	RIP	PA-90005 (B)	461	67.01	2.19	0.55	1.82	0.78	1.17	1.39	0.45	1.52	0.35	0.59	1.29	0.68	0.67	0.80	0.89	0.71	1.30	1.89	1.06	DQ661532
Gal2 POL23 M971	RIP	PA-90005 (B)	461	64.01	2.36	0.45	1.13	0.84	1.22	1.32	0.34	1.44	0.42	0.56	1.31	0.57	0.72	0.90	0.75	1.03	1.19	1.58	1.04	DQ661516
Gal2 POL23 M87	RIP	PA-90005 (B)	461	61.34	1.77	0.77	1.73	0.61	1.26	1.19	0.49	1.44	0.70	0.61	1.68	0.95	0.69	0.79	0.94	0.64	1.11	1.43	1.04	DQ661517
Average, C. reinhardtii genome			76,594	51.03	0.98	1.07	0.83	0.99	1.04	1.04	1.02	1.01	1.12	0.84	0.55	1.02	0.86	0.83	0.86	0.94	0.75	1.29	1.28	
Average, "normal" transgenes sequences			14,179	49.32	0.79	1.19	0.88	0.91	1.09	1.30	0.91	0.99	1.12	0.95	0.74	0.96	0.73	0.93	0.89	1.02	0.90	1.09	1.21	
Average, RIP-mutated transgenes sequences			19,210	57.62	2.30	0.44	1.93	0.78	1.05	1.19	0.31	1.66	0.58	0.68	1.35	0.55	0.85	0.99	1.00	0.55	1.06	1.57	0.88	

Table 2.3

Results of the Shimodaira-Hasegawa likelihood ratio test of the 20 sliding window consensus tree from the *Ccret2*POL2/3 dataset used to statitically test whether discordant topologies were generated along the length of the nucleotide sequence alignment.

Tree	-ln L	Diff -ln L	P
1	2799.32	287.52	0.0000*
2	2513.44	1.65	0.9539
3	2511.79	(best)	--
4	2512.59	0.79	0.9691
5	2607.40	95.61	0.1075
6	2635.69	123.89	0.0269*
7	2638.74	126.94	0.0240*
8	2760.59	248.80	0.0000*
9	3097.23	585.43	0.0000*
10	2730.17	218.37	0.0001*
11	2614.29	102.50	0.0688
12	2834.20	322.40	0.0000*
13	2647.30	135.50	0.0140*
14	2835.40	323.60	0.0000*
15	3130.95	619.15	0.0000*
16	2948.47	436.68	0.0000*
17	2795.08	283.29	0.0000*
18	2795.08	283.29	0.0000*
19	2809.97	298.17	0.0000*
20	2802.66	290.87	0.0000*

** $P < 0.05$*

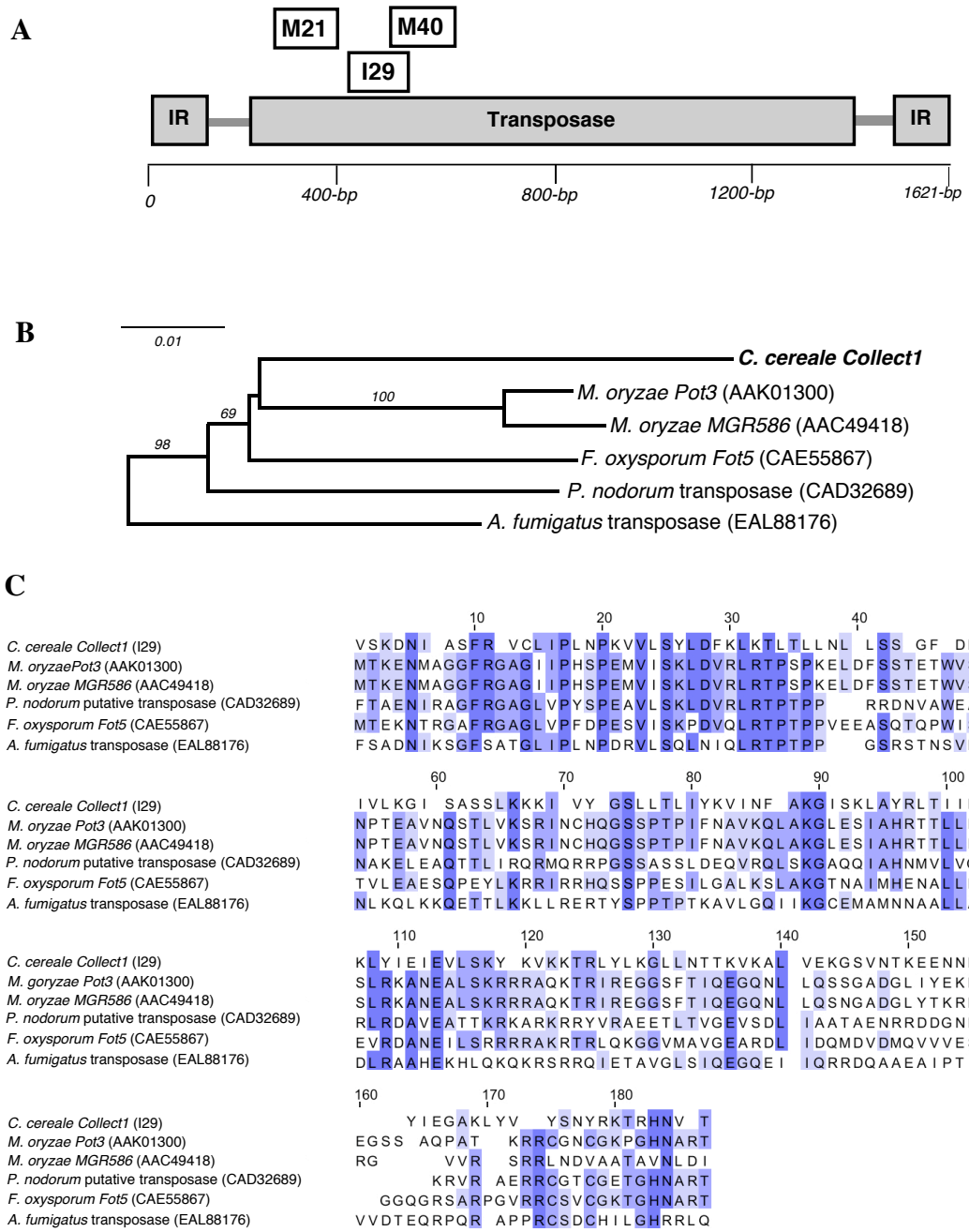


Figure 2.1

The *pogo* family DNA transposon *Collect1*. (A) Diagram of the *Pot3* transposon from *Magnaporthe oryzae* showing the relative positions of the I29, M21 and M40 elements from *Colletotrichum cereale* isolate PA-50005. (B) Phylogenetic tree inferred from the amino acid alignment. Posterior probabilities are shown above the branches. (C) Amino acid alignment with putative homologues; dashes represent gaps in the alignment. Shading represents levels of conservation (dark shading=highly conserved position, white=poorly conserved).

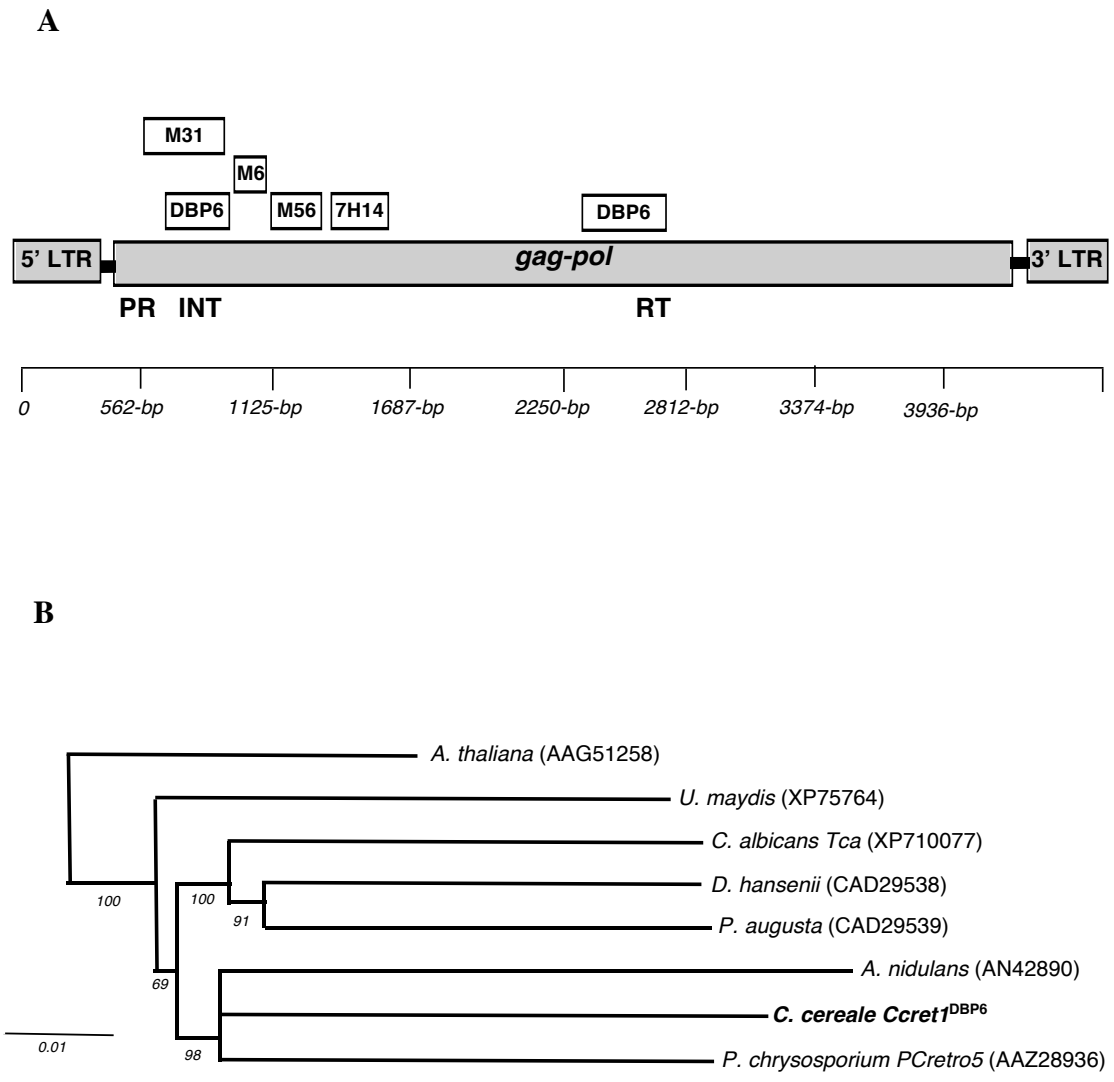


Figure 2.2 (a) and (b)

The *Pseudoviridae* family DNA retrotransposon *Ccret1*. (A) Diagram of the *PCretro5* retrotransposon (AAZ28936) from *Phanerochaete chrysosporium* showing the relative positions of *Ccret1* elements M6, M31, M56, 7H14 and DBP6 from *Colletotrichum cereale* isolate PA-50005. (B) Phylogenetic tree inferred from the amino acid alignment. Posterior probabilities are shown above the branches.

C

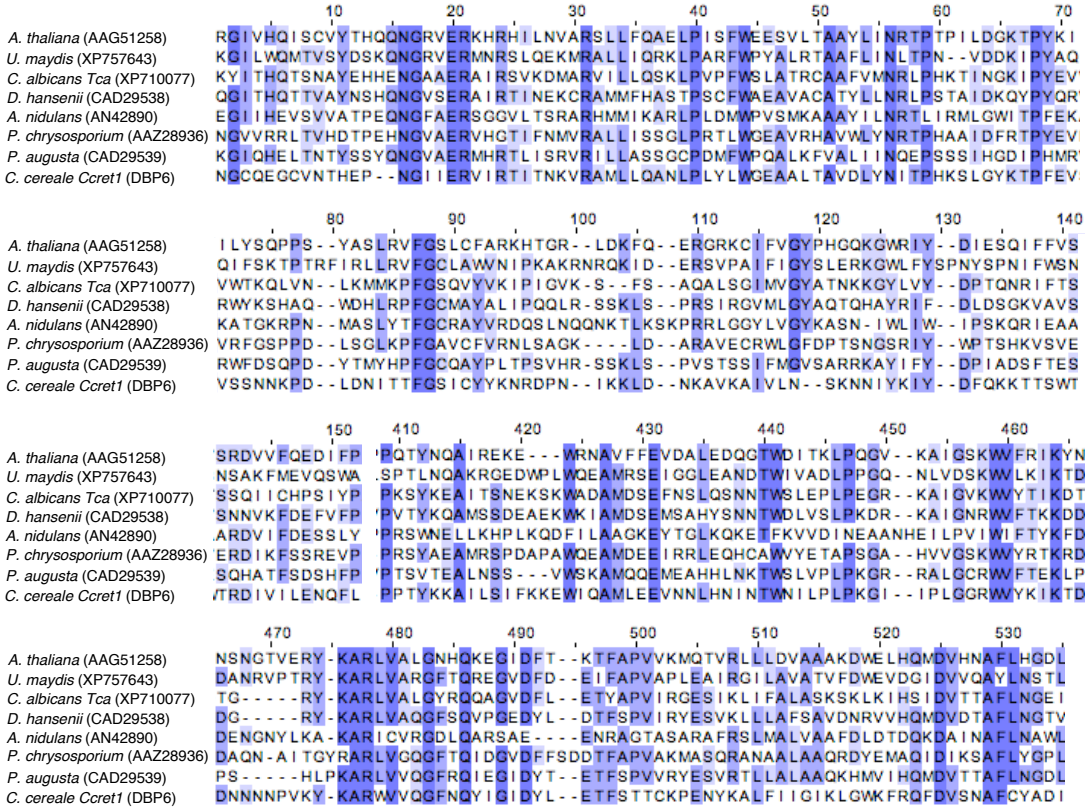
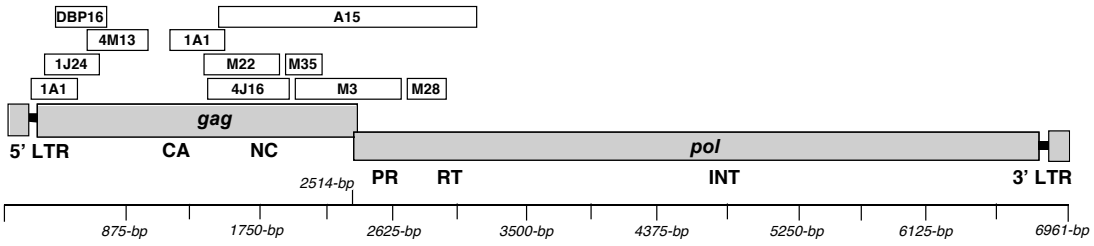


Figure 2.2 (c)

The *Pseudoviridae* family DNA retrotransposon *Ccret1*. (C) Partial amino acid alignment with DBP6 and its putative homologues; dashes represent gaps in the alignment. Only the relatively conserved 234 residues used for the phylogenetic analysis are shown. Shading represents levels of conservation (dark shading=highly conserved position, white=poorly conserved).

A



B

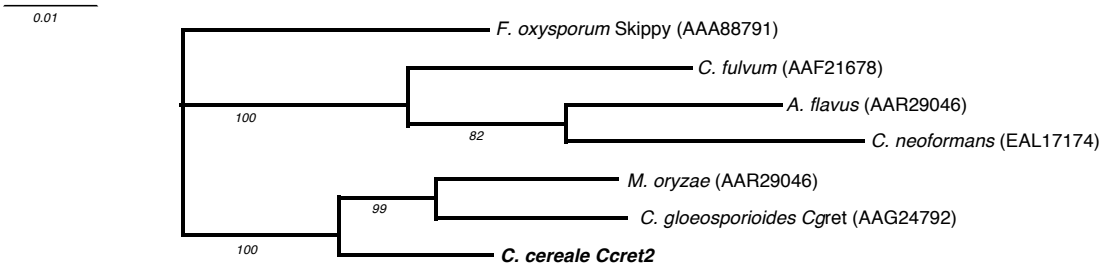


Figure 2.3 (a) and (b)

The *Metaviridae* family retrotransposon *Ccret2*. (A) Diagram of the retrotransposon *Cg*ret (AAG24792) from *Colletotrichum gloeosporioides* showing the relative positions of *Ccret2* elements A15, DBP16, 4J16, M3, M22, 1A1, 1J24, 4M13, M28 and M35. (B) Phylogenetic tree inferred from the amino acid alignment. Posterior probabilities are shown above the branches.

C

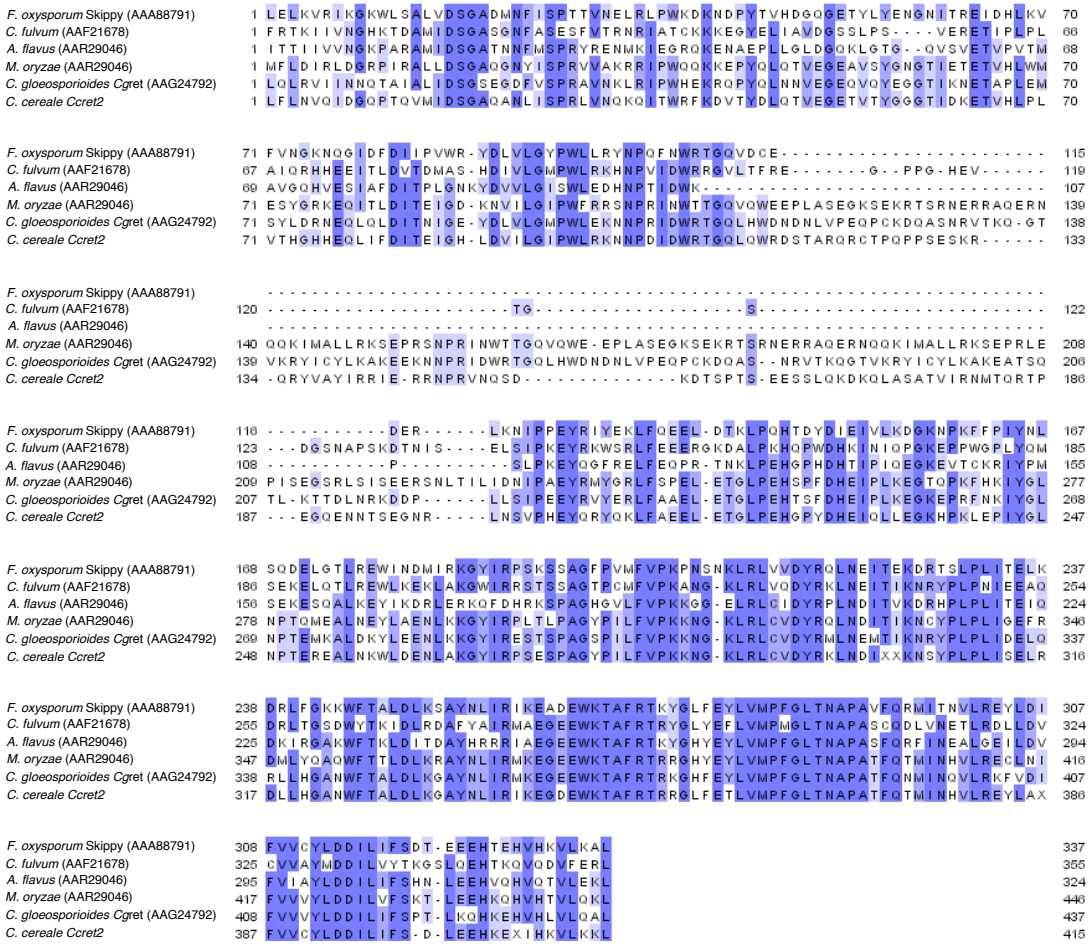


Figure 2.3 (c)

The *Metaviridae* family retrotransposon *Ccret2*. (C) Partial amino acid alignment with A15 and its putative homologues; dashes represent gaps in the alignment. Shading represents levels of conservation (dark shading=highly conserved position, white=poorly conserved).

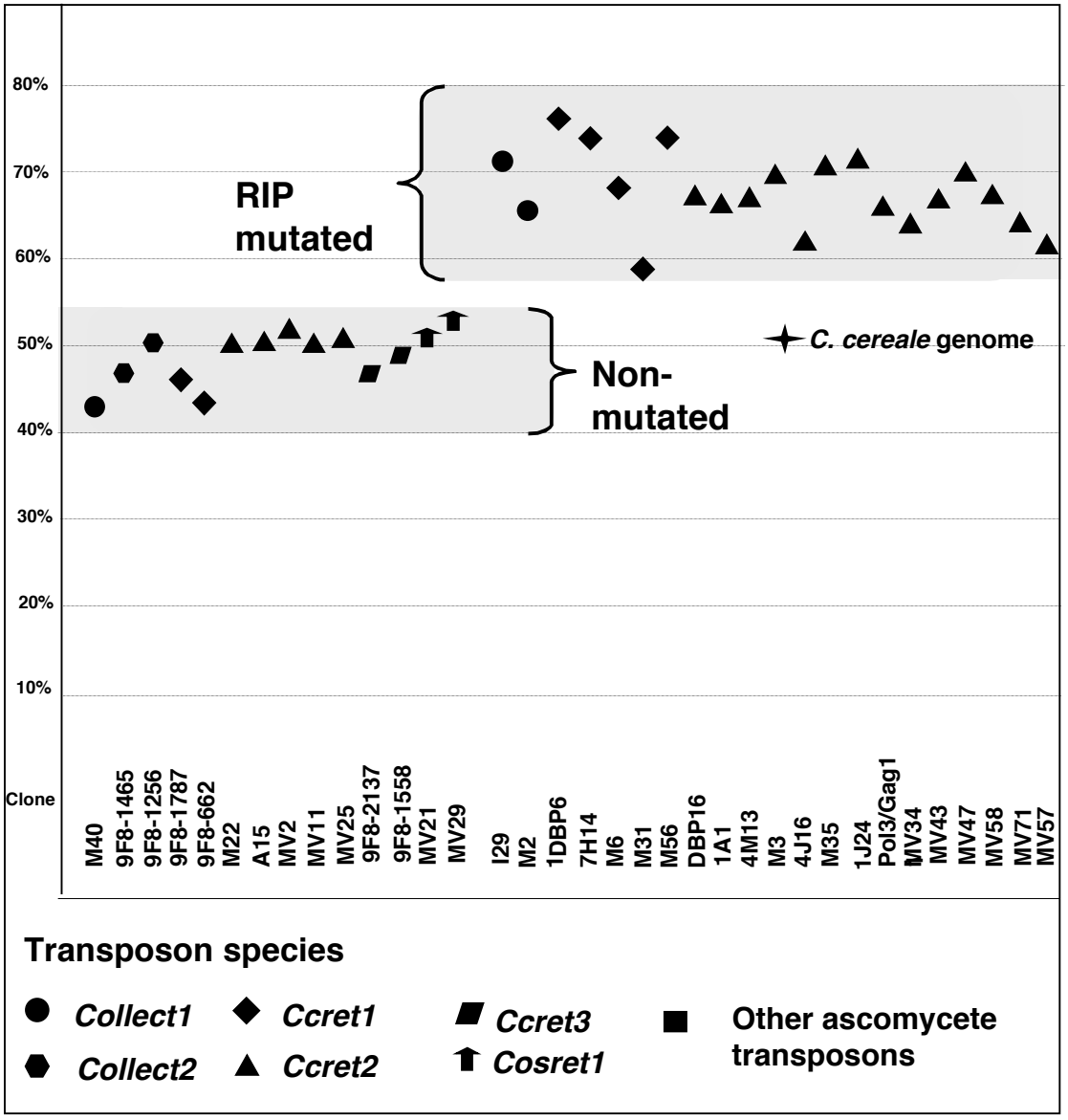


Figure 2.4a
The *Colletotrichum cereale* transposon nucleotide base composition (A+T%).

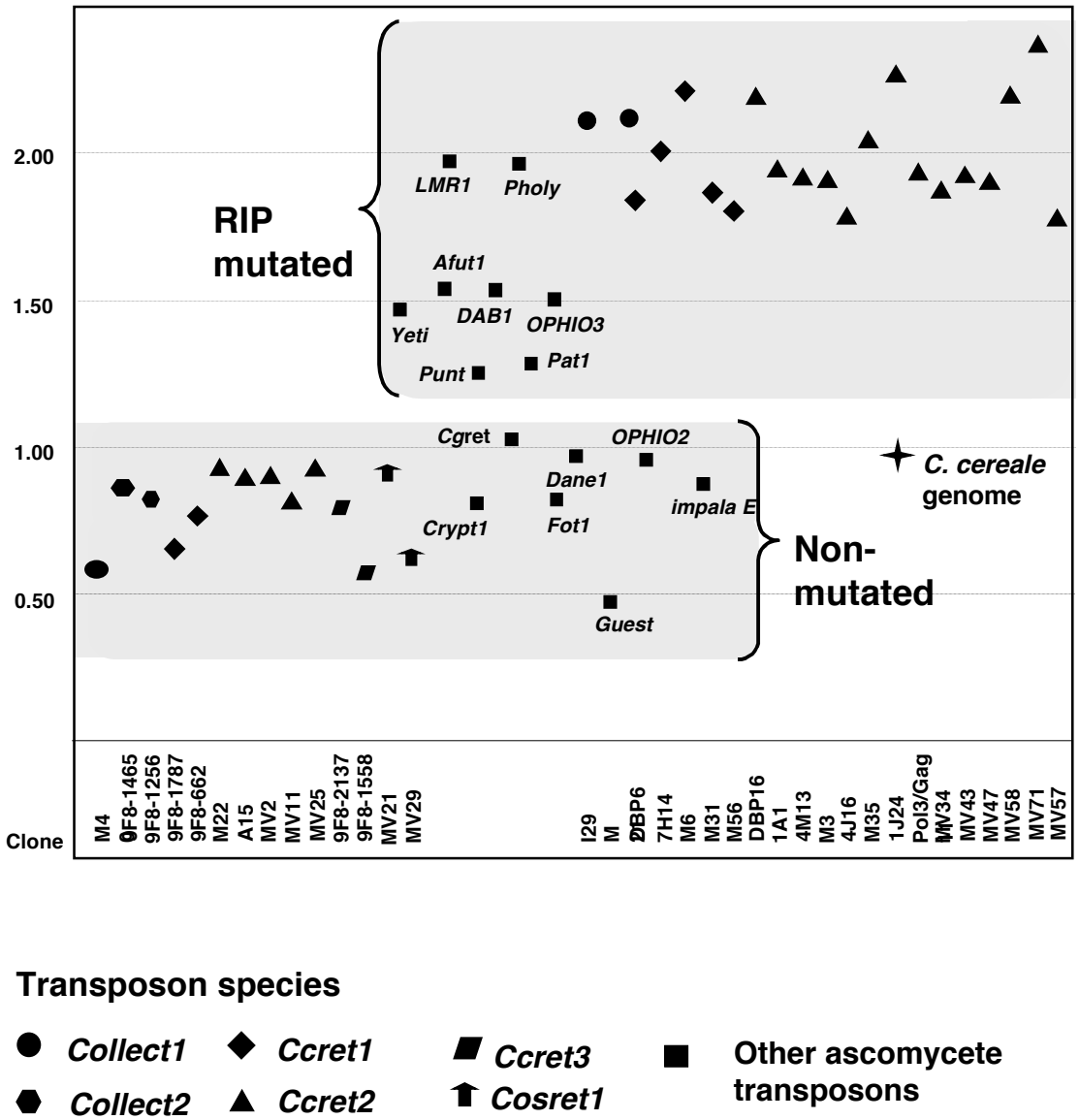


Figure 2.4b

Transposon TA/AT ratios for *C. cereale* and other ascomycete species. Accession numbers for additional RIPPED transposons: *Aspergillus nidulans Afut1*: L76086; *Fusarium oxysporum Fot1*_RIP: AF434909; *Leptosphaeria maculans Pholy*: AM084367; *L. maculans* LMR1: LM77515; *Neurospora crassa DAB1*: Y14976; *N. crassa Punt*: AF181821; *Ophiostoma ulmi OPHIO3*: DQ649005; *Podospora anserina Pat1*: AJ270953; *P. anserina Yeti*: AJ272171. Accession numbers for additional, non-mutated transposons: *A. nidulans Dane1*: AF295689; *C. gloeosporioides Cg*ret: AF264028; *Cryphonectria parasitica Crypt1*: AF283502; *F. oxysporum impala E*: AF363407; *F. oxysporum Fot1*: X64799; *N. crassa Guest*: AY197334; *O. ulmi OPHIO2*: DQ649004.

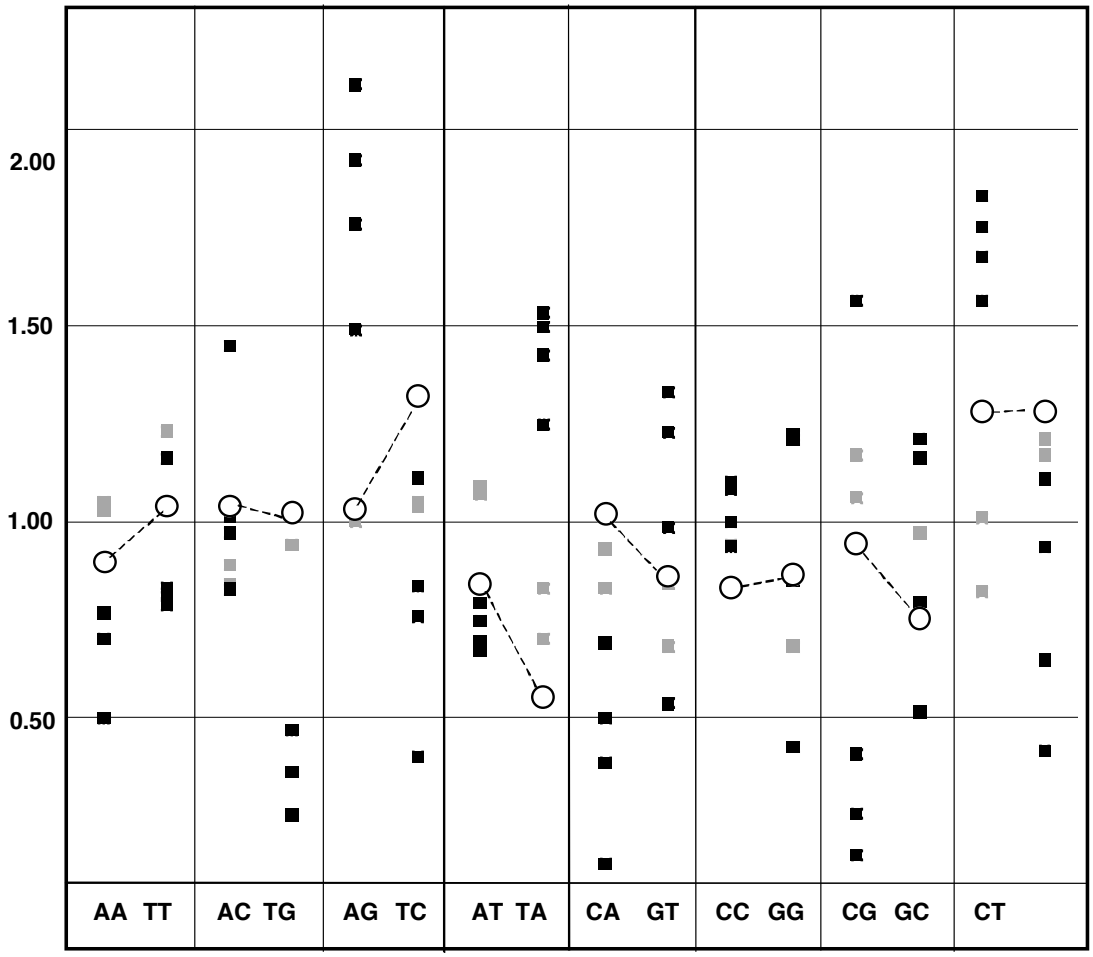


Figure 2.5 (a)
Frequencies (observed/ expected) of dinucleotide pairs in *Colletotrichum cereale* transposon sequences. (A) *Ccret1*

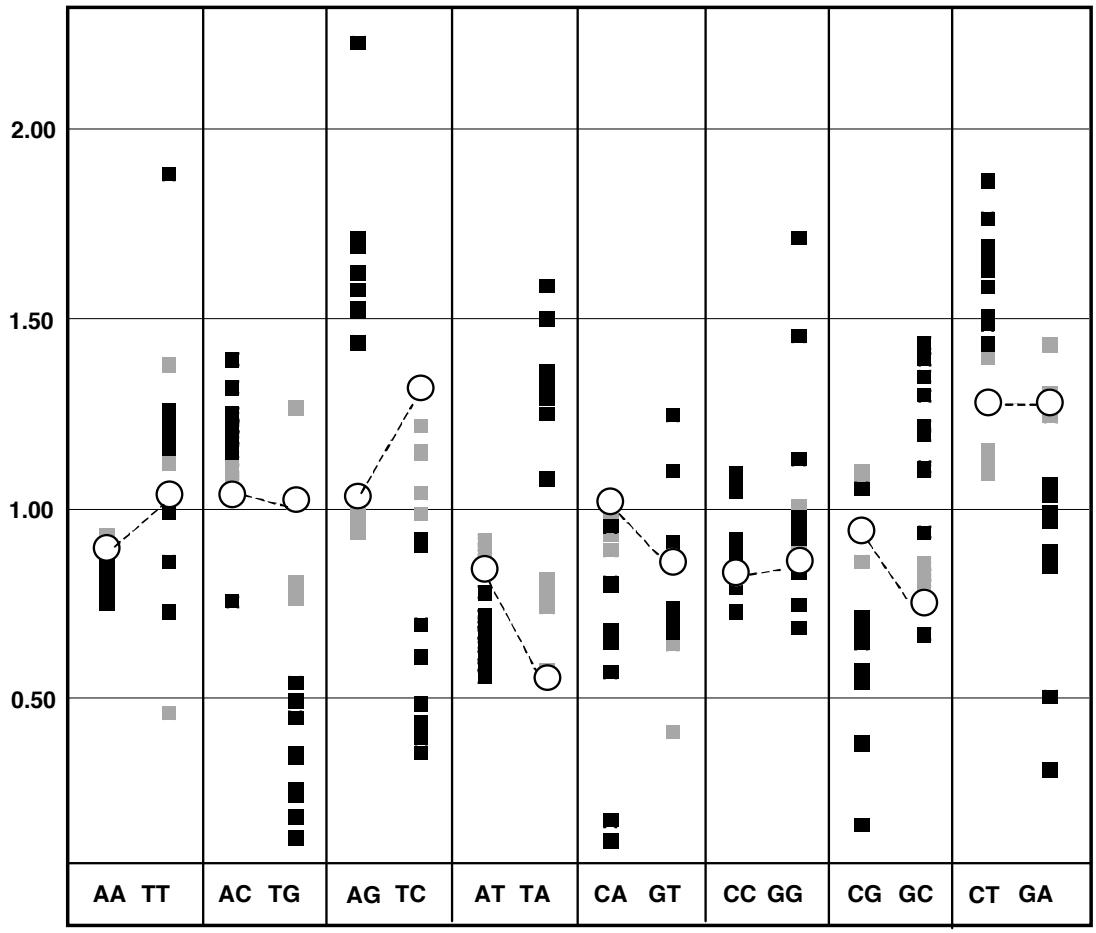


Figure 2.5 (b)
Frequencies (observed/ expected) of dinucleotide pairs in *Colletotrichum cereale* transposon sequences. ((B) *Ccret2*.

A.

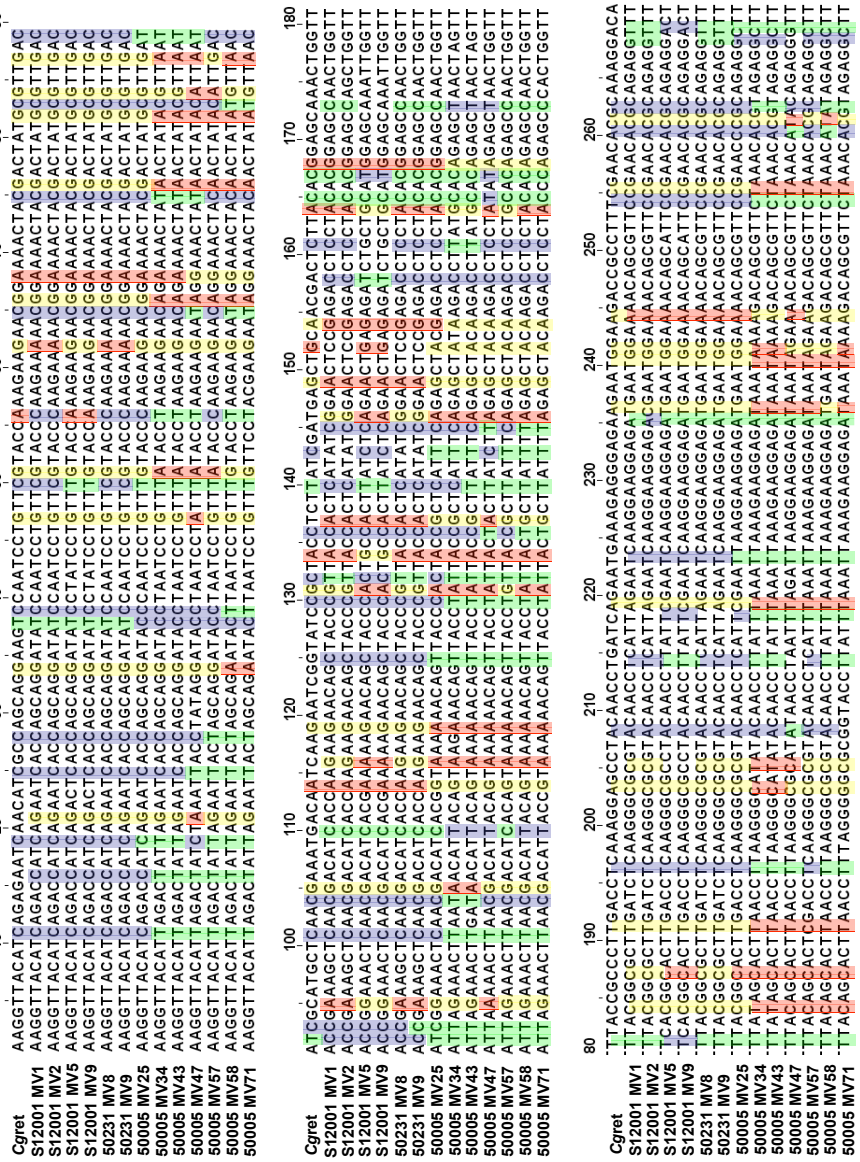


Figure 2.6 (a)
Ccret2$^{POL2/3}$ phylogenetic analysis. Partial alignment of the nucleotide dataset.

B.

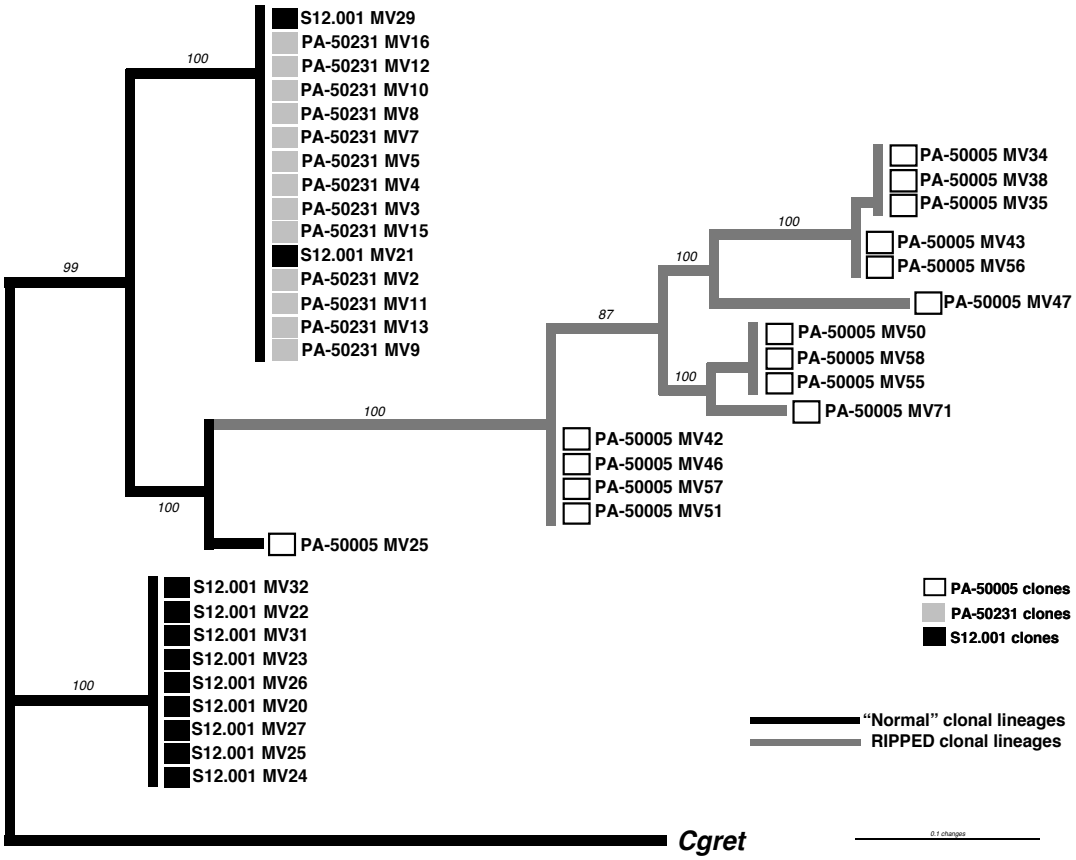


Figure 2.6 (b)

Ccret2$^{POL2/3}$ phylogenetic analysis. The phylogenetic tree was constructed from cloned PCR amplicons (540-bp) from the *pol* region of *Ccret2*^{A15} in three *Colletotrichum* lineages: PA-50231 (14 clones, *Colletotrichum cereale* clade A), PA-50005 (15 clones, *C. cereale* clade B), and S12.001 (11 clones, *C. sublineolum*); *Cg*ret from *C. gloeosporioides* serves as the outgroup. The phylogenetic tree was generated from 75% consensus of 43,601 trees estimated using Bayesian phylogenetic inference.

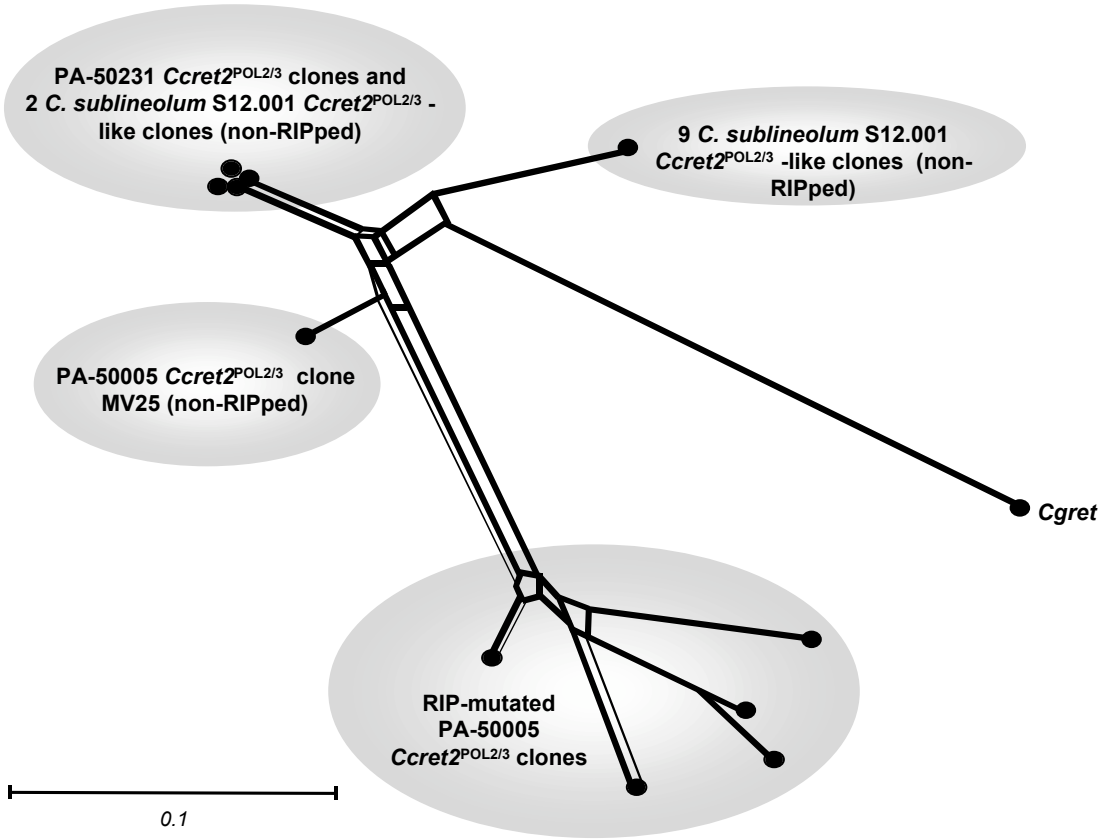


Figure 2.7

Ccret2$^{POL2/3}$ network using LogDet distance analysis shows the pattern of reticulate relationships between the cloned sequences from *Colletotrichum cereale* isolates PA-50231 (clade A), PA-50005 (clade B), and *C. sublineolum* isolate S12.001. The LTR-retrotransposon *Cg*ret from *C. gloeosporioides* is included as an outgroup

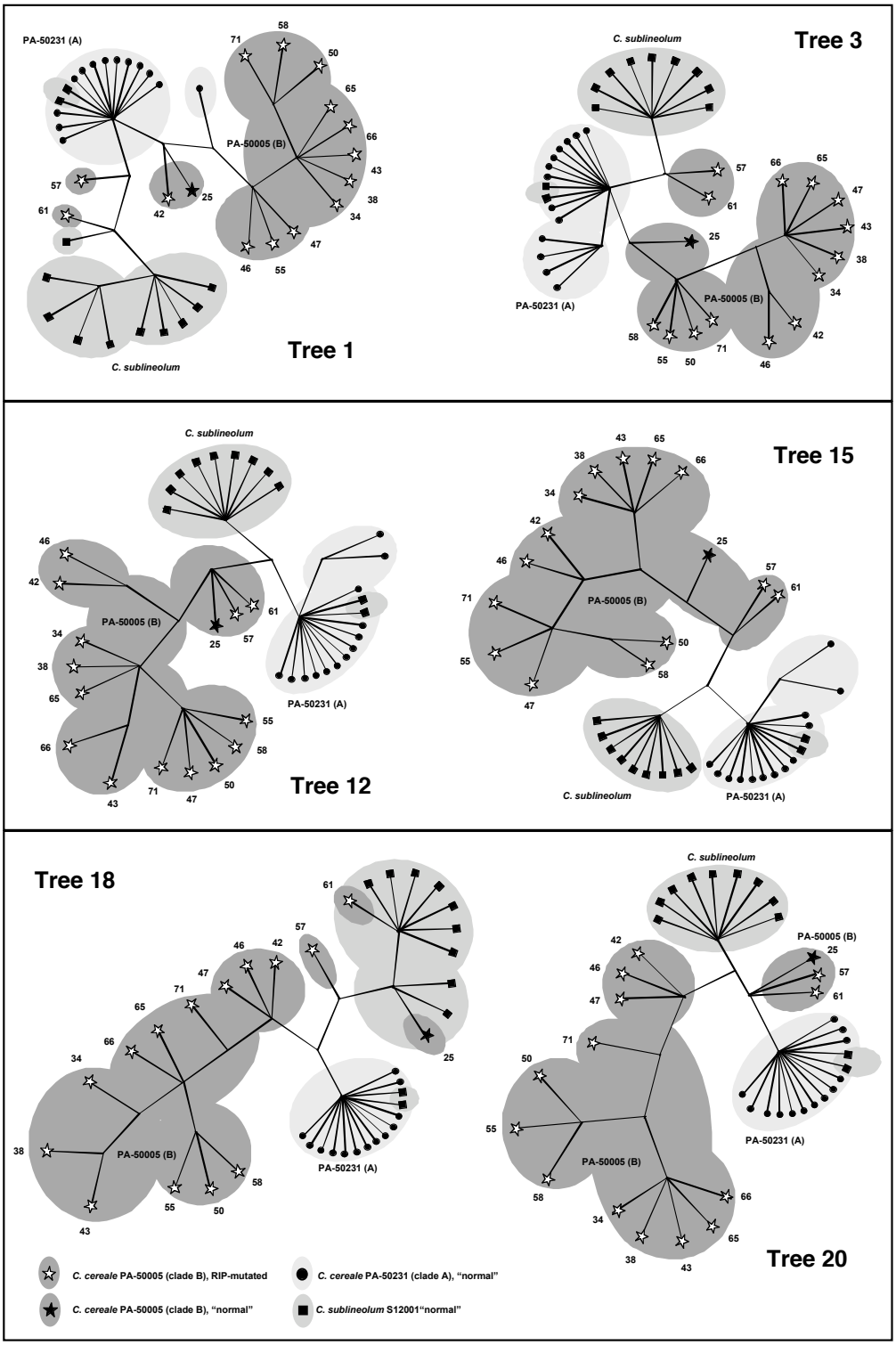


Figure 2.8
Representative sample of tree topologies recovered using the sliding window analysis of the *Ccret2*$^{POL2/3}$ dataset.

Chapter 3: Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles

ABSTRACT

Anthracnose disease of cool-season turfgrasses, caused by the fungus *Colletotrichum cereale*, has recently emerged as one of the most significant pathogens of *Poa annua*. Here we investigated the utility of four repetitive transposable elements as molecular markers for the analysis of *C. cereale* populations. Southern blot hybridization analysis revealed lineage-specific polymorphisms and distribution patterns for these transposons. Comparative phylogenetic analysis of three non-repetitive protein coding DNA sequences against the transposon RFLPs indicated that the transposon sequences have similar evolutionary histories to those found in the sampled *C. cereale* population, despite the alteration of several transposon copies by repeat-induced point (RIP) mutation. The variability and ubiquity of the *Ccret2*^{A15} transposon in *C. cereale* genomes suggests that this element could be utilized as a reliable DNA marker to discriminate between lineages of the fungus, identify hybrid genotypes and analyze genetic diversity in populations of this turfgrass pathogen.

3.1 Introduction

During the past decade the anamorphic fungus *Colletotrichum cereale* sensu lato Crouch, Clark and Hillman (formerly *C. graminicola* G.W. Wilson [Crouch *et al.*, 2006]) emerged from relative obscurity to become one of the most devastating pathogens of the

cool-season turfgrass *Poa annua*, causing epidemics of anthracnose disease in stands of this grasses maintained as golf course greens in North America (Smiley *et al.*, 2005) and the United Kingdom (Mann and Newell, 2005). For golf course superintendents, management of anthracnose is a challenging and expensive undertaking. Control of the disease is heavily reliant upon fungicide applications; however, resistance to benzimidazole, strobilurin, and sterol inhibitor fungicidal chemistries is an increasingly widespread phenomenon (Avila-Adame et al., 2003; Crouch et al., 2005; Wong and Midland, 2007; Wong et al., 2007; B.B. Clarke, unpublished data).

Because genetic variability between isolates of *C. cereale* may influence the trajectory of anthracnose disease of turfgrass, a comprehensive understanding of how *C. cereale* populations are organized and distributed across their geographic range could enhance the development and implementation of effective disease management strategies. At present only limited population level data, derived from randomly amplified polymorphic DNA (RAPD) or isozyme markers, is available for the fungus (Backman et al., 1999; Browning et al., 1999; Chen et al., 2002; Horvath and Vargas, 2004), although two major lineages, designated clades A and B, have been recognized based on intergenic transcribed spacer (ITS) nucleotide sequences (Crouch *et al.*, 2005) and a multiple gene genealogical approach (Crouch *et al.*, 2006). Currently there are few apparent biological patterns readily ascribable to this divergence and uncertainty exists as to whether the two groups are genetically isolated. *C. cereale* clades A and B are morphologically indistinguishable and have overlapping distributions; furthermore, each lineage includes a cohort of both disease-inducing isolates from turfgrass species and their non-pathogenic counterparts from cereal crops and natural grassland ecosystems (Crouch *et al.*, 2006; J.A. Crouch and B.I. Hillman, *unpublished data*).

The presence or absence of transposons at particular loci is a major contributor to RFLP variation in filamentous fungi. The primary objective of this research was to determine if repetitive transposable elements from the *C. cereale* genome could be developed as molecular markers to assess population structure and variability in the species. In the present study, we evaluated four elements representing three species of transposons (Crouch *et al.*, 2007) from *C. cereale* as molecular markers to examine population structure in this organism. Because of their ubiquitous and repetitive nature, molecular marker systems based on mobile transposable element polymorphisms have been employed for population-level analyses of numerous organisms, including several filamentous fungi (Diez *et al.*, 2003; Farman *et al.*, 1996; Girard and Freeling, 1999; Kohn *et al.*, 1991; Linder-Basso *et al.*, 2001; Milgroom *et al.*, 1992). The presence of a transposon at a genomic locus is typically a good indicator of identity by descent, while the absence of an element at a site is recognized as the ancestral state. Transposon insertional restriction fragment length polymorphism (RFLP) data can be relatively free of homoplasic data that might be inconsistent with an organism's evolutionary history, since the independent insertion of two different transposon copies at the exact same location on a chromosome is extremely unlikely. The parallel loss of transposon copies through excision or homologous recombination may be problematic, however (Carbone *et al.*, 1999), and alteration by repeat-induced point (RIP) mutation of transposons may complicate the evolutionary signal (Crouch *et al.*, 2007). Although base substitutions in the restriction enzyme recognition sequence can theoretically generate non-homologous bands of identical size, there is little likelihood of this type of convergence in groups of closely related individuals where evolutionary rates are low; in particular, parallel independent gains of restriction sites occur with only a small probability (Nei and Tajima,

1983; Nei and Tajima, 1985; Upholt, 1977)

The objectives of this study were to determine if transposon RFLP markers supported the separation of *C. cereale* into two distinct lineages as previously described (Crouch *et al*, 2005; Crouch *et al.*, 2006) and to examine whether these markers offered any advantages over nucleotide sequence data in discerning structure in *C. cereale* populations. In particular, we considered to what extent these transposons could extend our understanding of how the major *C. cereale* lineages have evolved.

3.2 Materials and methods

3.2.1 Fungal cultures

Twenty-one single spore cultures of *C. cereale* were isolated from diseased *Poa annua* on 11 golf course greens located within a 100 km radius in Pennsylvania (Fig. 1, Table 1) and cultured as previously described (Crouch *et al.*, 2006). Isolates of *C. graminicola* from *Zea mays*, *C. sublineolum* from *Sorghum bicolor* and *C. falcatum* from *Saccharum officinarum* were used for outgroup comparisons.

3.2.2 Restriction fragment length polymorphism (RFLP) analyses

Genomic DNA was isolated from mycelia using a standard phenol:chloroform extraction protocol (Sambrook and Russell, 2001). *Hin*dIII-digested genomic DNA was size fractionated by gel electrophoresis for 18 hours at 45v in 1x TBE buffer, then visualized using ethidium bromide staining. Southern blots for RFLP analysis were prepared by transferring the DNA to Zeta-Probe membranes (Bio-Rad, Hercules, CA) using a Posiblot Pressure Blotter (Strategene, La Jolla, CA) at 75 mm Hg. 500 ng of PCR

amplicon from each of the four individual transposon sequences (Table 2) was radiolabeled with [α^{32}P]dCTP (MP Biomedicals, Irvine, CA) using the Random Primers DNA Labeling System (Invitrogen Corp., Carlsbad, CA). Hybridizations were performed as previously described (Crouch *et al.*, 2007). Hybridized membranes were exposed to autoradiography film (Lab Scientific, Inc., Livingston, NJ) in the presence of an intensifying screen and incubated at -70°C for 24 to 48 hours before development. RFLP banding patterns of four sequences from three transposon species (*Collect1*^{I29}, *Ccret1*^{DBP6}, *Ccret2*^{DBP16} and *Ccret2*^{A15}) were evaluated.

Because the retrotransposon sequences *Ccret1*^{DBP6} and *Ccret2*^{A15} were identified in all of the *C. cereale* isolates sampled, the RFLP patterns from these elements were used to discern patterns of population subdivision. Bands on the autoradiograms were scored visually as either present or absent and coded as binary data. The datasets were analyzed to identify population groupings using the Bayesian Monte Carlo Markov chain-based clustering program Structure 2.1 (Falush *et al.*, 2003; Pritchard *et al.*, 2000) for 1,000,000 repetitions each, with the first 20,000 discarded as burn-in. These analyses were run using the admixture model and correlated allele frequencies between populations, which is considered the best strategy for the detection of subtle differences in population structure (Falush *et al.*, 2003). The degree of α admixture was empirically derived from the data, and the distribution of allelic frequencies λ was set to 1 (Falush *et al.*, 2003). 20 runs were performed for K=1 through 10 (where K=the maximum number of populations).

3.2.3 Phylogenetic analyses

Phylogenetic analysis was performed using three nuclear loci previously shown capable of differentiating between the two major lineages of *C. cereale*, with PCR amplified fragments of the *ITS1/5.8S/ITS2* ribosomal DNA (ITS), the HMG-box of the *Mat-1-2* mating idiomorph (HMG), and the manganese superoxide dismutase (*Sod2*) genes used to generate nucleotide sequence data as previously described (Crouch *et al.*, 2006). The sister species of *C. cereale* -- *C. sublineolum* and *C. falcatum* -- along with the more distantly related species *C. graminicola* (Crouch *et al.*, 2006), J.A. Crouch and B.I. Hillman, *unpublished data*) were included as outgroup taxa. Multiple sequence alignments were constructed using Clustal W (Thompson *et al.*, 1994) as launched in MegAlign (DNASTAR, Inc., Madison, WI), and manually adjusted to exclude gaps and ambiguously aligned regions from the dataset. Tree topologies were estimated from the combined multilocus nucleotide sequence dataset in MrBayes v.3.1 (Huelsenbeck and Ronquist, 2003) by performing two simultaneous runs of one cold and three heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC) for 40,000,000 generations and sampling trees every 500 generations. Each individual gene region was partitioned in the analysis, and a general evolutionary model for each partition was incorporated as selected using the program ModelTest v.3.06 (Posada and Crandall, 1998) (ITS model: TrNef+G, A→G 1.5282, C→T 3.9607; α=0.1317; HMG model: HKY+G, A=0.2654, C=0.2953, G=0.2622, T=0.1770; Ti/Tv=1.2783; α=1.50421; *Sod2* model: TrN+I, A=0.2477, C=0.3035, G=0.2612, T=0.1876; A→G 5.0186, C→T 4.8842, G→T 1.0; Pinv=0.4998; equal rates for all sites). Run diagnostics were performed every 1000th generation, with the first 25% of the output discarded as burn-in. To confirm that the two separate runs converged to a stationary distribution, the average standard deviation of split frequencies, the potential scale reduction factor convergence

diagnostic, and the plot of generation versus log likelihood were each examined for values and distributions characteristic of samples drawn from a posterior probability distribution. Trees sampled from the posterior distribution were imported into PAUP* v.4.0b10 (Swofford, 2000) and used to construct 75% majority-rule consensus trees.

3.2.4 Nucleotide sequences

All new sequences generated by this study have been deposited in the NCBI GenBank database (accession numbers **DQ663514**–**DQ663534**).

3.3 Results

3.3.1 Phylogenetic assessment of populations using sequence data

The fungal specimens included in this study, although drawn from a geographically limited region in Pennsylvania, represent both of the major *C. cereale* evolutionary lineages, clades A and B, allowing us to test whether the multilocus RFLP banding patterns of four sequences from three transposon species (*Collect1*^{I29}, *Ccret1*^{DBP6}, *Ccret2*^{DBP16} and *Ccret2*^{A15}) could be used to distinguish the major lineages in this species, even on a relatively fine-scale. Three of the probes -- *Collect1*^{I29}, *Ccret1*^{DBP6}, *Ccret2*^{DBP16} -- have been altered in the past through repeat-induced point (RIP) mutation, a genome defense system deployed by filamentous fungi that produces C→T and G →A transitions in repetitive DNA (Cambareri et al., 1989), including transposable elements. To evaluate the transposon-based population hypotheses, a strict consensus tree of 21 *C. cereale* isolates was constructed from 33,206 trees using Bayesian estimates from the combined ITS/HMG/*Sod2* dataset (Fig. 2). Both lineages were represented in the tree topology and

supported by posterior probabilities of 100, with 13 isolates from *C. cereale* clade A and 8 isolates from clade B. Two of the geographic locations contained isolates from each of the two clades.

3.3.2 Limited distribution of the TE sequences *Collect1*^{I29} and *Ccret2*^{DBP16}

The TE markers *Collect1*^{I29} and *Ccret2*^{DBP16} produced fingerprint profiles largely consistent with the phylogenetic groups and confirmed the repetitive nature of the transposon sequences when hybridized against the restricted DNA blots. Isolates phylogenetically characterized as *C. cereale* clade B resulted in ~25 hybridizing bands on the autoradiograms (Fig. 3) with little polymorphism observed between the individual isolates. In contrast, all clade A isolates except PA-50183 were devoid of the *Collect1*^{I29} and *Ccret2*^{DBP16} sequences, as were the outgroup samples of *C. graminicola* and *C. sublineolum*. PCR amplification using several alternate primer pairs from *Collect1*^{I29} and *Ccret2*^{DBP16} recovered the same pattern of presence/absence, failing to yield a product in clade A isolates even under conditions of low stringency (data not shown). The presence of these two elements in the genomes of *C. cereale* clade B and not in clade A is consistent with the fact that both of these transposon sequences are extensively RIP mutated, a process that has not been observed for clade A strains of the fungus (Crouch *et al*, 2007). But the PCR-based identification of *Ccret2*^{DBP16} from two of the three *C. falcatum* outgroup strains (data not shown) suggests that this RIP-mutated element was already present in the common ancestor of *C. falcatum* and *C. cereale* and was subsequently lost from *C. cereale* clade A after its divergence from clade B (Fig. 4).

3.3.3 The retrotransposons *Ccret1*^{DBP6} and *Ccret2*^{A15} are found in both *C. cereale* lineages

In contrast to the limited distribution of *Collect1*^{I29} and *Ccret2*^{DBP16} within the species, Southern blot analysis of the *C. cereale* population (Fig. 3) using the RIP-mutated *Ccret1*^{DBP6} probe revealed the presence of this retrotransposon in both of the major *C. cereale* lineages, although PCR amplification using a range of high and low stringency conditions and primer pairs demonstrated that *Ccret1*^{DBP6} was absent from the DNA of *C. graminicola*, *C. sublineolum* and *C. falcatum* (data not shown). Each of the *C. cereale* clades exhibited visually distinct banding patterns. Clade B isolates yielded between 9-15 *Ccret1*^{DBP6} bands ranging in size from ~0.5 kb to 9 kb, but, with the exception of isolate PA-50183, the clade A isolates faintly hybridized at only one or two restriction fragments. Low copy number of *Ccret1*^{DBP6} in the genome of clade A isolates was anticipated, since analysis of the element from a cosmid library found that this retrotransposon is present only as two unmutated copies at a single genomic locus in clade A isolate NJ-6340 (Crouch *et al.*, 2007). The observed faint hybridization to the RIP-mutated probe sequence was similarly predicted from the cosmid sequence data, since this transposon was not found to be RIP-mutated in clade A (Crouch *et al.*, 2007). All *C. cereale* isolates shared the 1 kb *Ccret1*^{DBP6} band, indicating that this is likely the ancestral locus of *Ccret1*^{DBP6} and that subsequent amplification and RIP-mutation of this retrotransposon occurred only after the divergence of clades A & B (Fig. 4).

The *Ccret2*^{A15} retrotransposon sequence was the only transposon used as a probe in this study that was not RIP-altered, although in clade B strains of the fungus, this

element can be present as both RIPped and non-RIPped variants within a single genome (Crouch *et al.*, 2007). Of the four sequences evaluated, *Ccret2*^{A15} was the only transposon that produced a polymorphic RFLP banding pattern (Fig. 3). Like the other three transposon probes, the *Ccret2*^{A15} marker produced a visually distinctive banding pattern clearly differentiating between isolates belonging to phylogenetic clades A and B. Likewise, clade A isolate PA-50183 exhibited the clade B-like fingerprint rather than the clade A-like pattern predicted by phylogenetic affiliation. PCR amplification identified *Ccret2*^{A15} from one of the two *C. sublineolum* isolates and all three of the *C. falcatum* isolates; however, it was absent from the more distantly related *C. graminicola*, suggesting that this transposon sequence was present in the common ancestor of *C. cereale*, *C. sublineolum* and *C. falcatum* (Fig. 4).

3.3.4 Estimates of population subdivision using the retrotransposon RFLP datasets

Since the *Ccret1*^{DBP6} and *Ccret2*^{A15} sequences were present in all of the *C. cereale* isolates sampled for this study, binary datasets were generated by coding the banding patterns produced by these elements as either present or absent to evaluate population sub-division. We first used the binary datasets to determine if the retrotransposon distribution within the genome was congruent with the HMG/ITS/*Sod2* evolutionary hypothesis. Consistent with the phylogenetic tree topology and the visual estimations made from the autoradiograms, two distinct populations, corresponding to Clades A and B, were inferred from the RFLP datasets using the Bayesian clustering method implemented in the program Structure (Pritchard *et al.*, 2000).

3.4 Discussion

Consistent with the multi-locus phylogenetic tree topology (Fig. 2), all four transposon RFLP fingerprint patterns recovered the division of *C. cereale* into two main lineages as previously established for the species (Crouch et al., 2006), either through distinct banding patterns or by their presence or absence. The only inconsistency observed between the nucleotide sequence data set and the transposon RFLPs was the manifestation of Clade B-like banding patterns for the Clade A isolate PA-50183 by all four transposon markers (Fig. 4), suggesting that this isolate might be a hybrid between the two lineages. Despite the potential for RIP-induced homoplasy in these analyses (Crouch *et al.*, 2007), our data showed the *C. cereale* transposon RFLP signal in these analyses to be largely congruent with the non-TE datasets, with both the RFLPs and sequence analysis of three protein coding genes yielding the same general conclusions. Although none of the RFLP fingerprints predicted any further population substructure beyond the two main lineages, this is likely a reflection of the small, geographically limited sample size evaluated in this study rather than a lack of sensitivity on the part of the markers. As the purpose of this study was to determine whether transposon RFLP patterns are suitable molecular markers rather than draw conclusions about the genetic makeup of populations, further study will be required to make this determination.

The interspecific distribution and intraspecific polymorphic banding patterns demonstrated that of the four markers evaluated, *Ccret2*^{A15} sequence has the potential to serve as an effective RFLP marker for future population analysis of *C. cereale* and might even be adopted for use in populations of the closely related, economically important

plant pathogens *C. sublineolum* and *C. falcatum*. *Ccret2*^{A15} is polymorphic and was present in all *C. cereale* isolates sampled in this study; additionally, PCR-based screening shows that this transposon is widely distributed across the geographic range for this species and is present in both turfgrass pathogenic strains as well as *C. cereale* isolated from prairie, forage and cereal crops (J.A. Crouch and B.I. Hillman, unpublished data). In contrast, while any of the other three transposons surveyed in this work – *Collect1*^{I29}, *Ccret1*^{DBP6} and *Ccret2*^{DBP16} – might, in theory, be employed to evaluate populations of *C. cereale* Clade B given the polymorphic banding patterns shown by the group, the high level of RIP mutation that characterizes these elements renders the use of these transposons as RFLP markers potentially problematic (Crouch *et al.*, 2007). Under normal circumstances, although base substitutions in a restriction enzyme recognition sequence can theoretically generate non-homologous bands of identical size, there is little likelihood of this type of convergence in groups of closely related individuals where evolutionary rates are low; in particular, parallel independent gains of restriction sites with 6-base recognition sequences are have been found to occur with only a small probability (Nei and Tajima, 1983; Nei and Tajima, 1985; Upholt, 1977). But for RIPped transposons, restriction sites are more rapidly gained or lost, since overall nucleotide composition and dinucleotide patterns are skewed, often occurring at a range of different levels contingent upon how many rounds of RIP-mutation have acted on a given element. Thus, because RIP mutation has been found to act on these transposons, we cannot exclude the possibility that the different allelic states (+/-) observed at each locus are merely artifacts of RIP alterations rather than accurately reflecting common descent. For these reasons, for *C. cereale* clade B and other fungi where there is evidence of RIP mutation, transposon RFLP datasets should be regarded as potentially homoplasic unless there is independently

derived support for the interpretation of homology. In the present study, however, the agreement between transposon RFLP data and the three independent protein coding genes attest to the consistency of the RFLP data in this sampled population and suggest that the *Ccret2*^{A15} transposon-based marker can serve as a valuable tool in future population studies of *C. cereale*.

3.5 References

Avila-Adame, C., G. Olaya, and W. Koller. 2003. Characterization of *Colletotrichum graminicola* isolates resistant to strobulurin-related QoI fungicides. *Plant Dis.* **87**:1426-1432.

Backman, P.A., P.J. Landschoot, and D.R. Huff. 1999. Variation in pathogenicity, morphology, and RAPD marker profiles in *Colletotrichum graminicola* from turfgrasses. *Crop Sci.* **39:**1129-1135.

Browning, M., L.V. Rowley, P. Zeng, J.M. Chandlee, and N. Jackson. 1999. Morphological, pathogenic, and genetic comparisons of *Colletotrichum graminicola* isolates from *Poaceae*. *Plant Dis.* **83**:286-292.

Cambareri, E.B., B.C. Jensen, E. Schabtach, and E.U. Selker. 1989. Repeat-induced G-C to A-T mutations in *Neurospora*. *Science* **244**:1571-5.

Carbone, I., J. Anderson, and L. Kohn. 1999. Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. *Evolution* **53**:11-21.

Chen, F., P.H. Goodwin, A. Khan, and T. Hsiang. 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris*. *Can. J. Microbiol.* **48**:427-36.

Crouch, J.A., B.B. Clarke, and B.I. Hillman. 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. *Int. Turfgrass Soc. Res. J.* **10**:186-195.

Crouch, J.A., B.B. Clarke, and B.I. Hillman. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. *Phytopathol.* **96**:46-60.

Crouch, J.A., B.M. Glasheen, M.A. Giunta, B.B. Clarke, and B.I. Hillman. 2007. The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. *Fungal Genet. Biol.* **45**:190–206.

Diez, J., T. Beguiristain, F. Le Tacon, J.M. Casacuberta, and D. Tagu. 2003. Identification of Ty1-copia retrotransposons in three ectomycorrhizal basidiomycetes: Evolutionary relationships and their use as molecular markers. *Curr. Genet.* **43**:34-44.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**:1567-1587.

Farman, M.L., S. Taura, and S.A. Leong. 1996. The *Magnaporthe grisea* DNA fingerprinting probe MGR586 contains the 3' end of an inverted repeat transposon. *Mol. Gen. Genet.* **251**:675-81.

Girard, L., and M. Freeling. 1999. Regulation changes as a consequence of transposon insertion. *Dev. Genetics* **25**:291-296.

Horvath, B.J., and J.M. Vargas. 2004. Genetic variation among *Colletotrichum graminicola* isolates from four hosts using isozyme analysis. *Plant Dis.* **88**:402-406.

Huelsenbeck, J.P., and F. Ronquist. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572-1574.

Kohn, L.M., E. Stasovski, I. Carbone, J. Royer, and J.B. Anderson. 1991. Mycelial incompatibility and molecular markers identify genetic variability in field populations of *Sclerotinia sclerotiorum*. *Phytopathol.* **81**:480-485.

Linder-Basso, D., R. Foglia, P. Zhu, and B.I. Hillman. 2001. *Crypt1,* an active *Ac*-like transposon from the chestnut blight fungus, *Cryphonectria parasitica*. *Mol. Gen. Genet.* **I265**:730-738.

Mann, R.L., and A.J. Newell. 2005. A survey to determine the incidence and severity of pests and diseases on golf course putting greens in England, Ireland, Scotland, and Wales. *Intl. Turfgrass Soc. Res. J.* **10**:224-229.

Milgroom, M.G., S.E. Lipari, and W.A. Powell. 1992. DNA fingerprinting and analysis of population structures of the chestnut blight fungus, *Cryphonectria parasitica*. *Genetics* **131**:297-306.

Nei, M., and F. Tajima. 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. *Genetics* **105**:207-217.

Nei, M., and F. Tajima. 1985. Evolutionary change of restriction cleavage sites and phylogenetic inference for man and apes. *Mol. Biol. Evol.* **2**:189-205.

Posada, D., and K.A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14**:817-818.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics 155*:945-959.

Sambrook, J., and D. Russell. 2001. Molecular Cloning: A Laboratory Manual. 3nd edition ed. Cold Spring Harbor Laboratory, New York.

Smiley, R.W., P.H. Dernoeden, and B.B. Clarke. 2005. Compendium of Turfgrass Diseases. 3rd ed. APS Press, St. Paul.

Swofford, D.L. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. *Nucleic Acids Res.* **22**:4673-4680.

Upholt, W.B. 1977. Estimation of DNA sequence divergence from comparison of restriction endonuclease digests. *Nucleic Acids Res.* **4**:1257-1265.

Vaillancourt, L., M. Du, J. Wang, J. Rollins, and R. Hanau. 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. *Mycologia* **92**:430-435.

White, T.J., T. Bruns, S. Lee, and J.L. Taylor. 1991. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p. 315-322, *In* M. A. Innis, et al., eds. PCR Protocols, a Guide to Methods and Applications. Academic Press, New York.

Wong, F.P., and S.L. Midland. 2007. Sensitivity distributions of California populations of *Colletotrichum cereale* to four sterol demethylation inhibitor fungicides: Propiconazole, myclobutanil, tebuconazole, and triadimefon. *Plant Dis*. **91**:1547-1555.

Wong, F.P., S.L. Midland, and K.A. de la Cerda. 2007. Occurrence and distribution of QoI-Resistant Isolates of *Colletotrichum cereale* from annual bluegrass in California. *Plant Dis*. **91**:1531-1546.

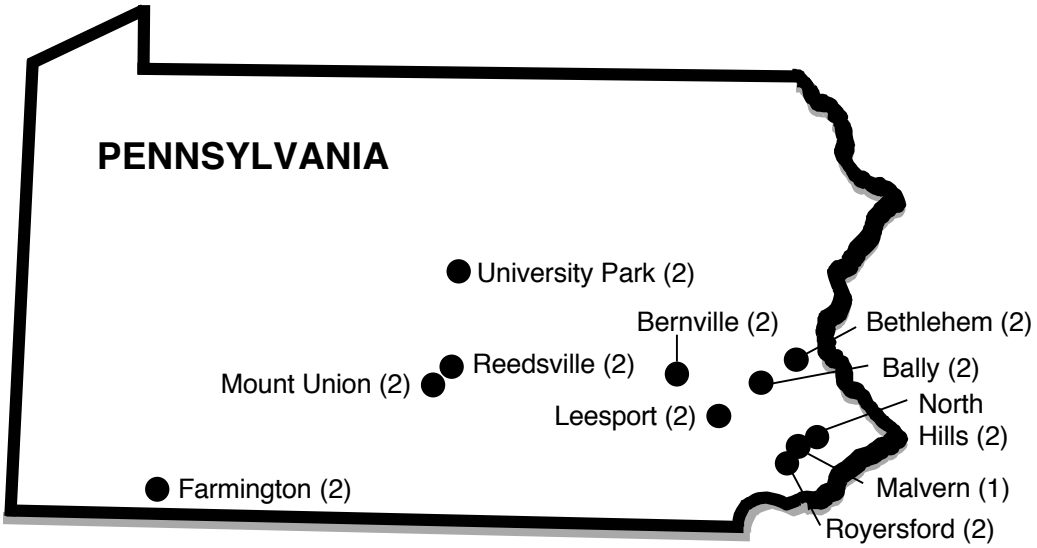


Figure 3.1
Map of Pennsylvania, illustrating the origination of the *Colletotrichum cereale* isolates used in this study. The number of isolates from each location is listed in parentheses after the location name.

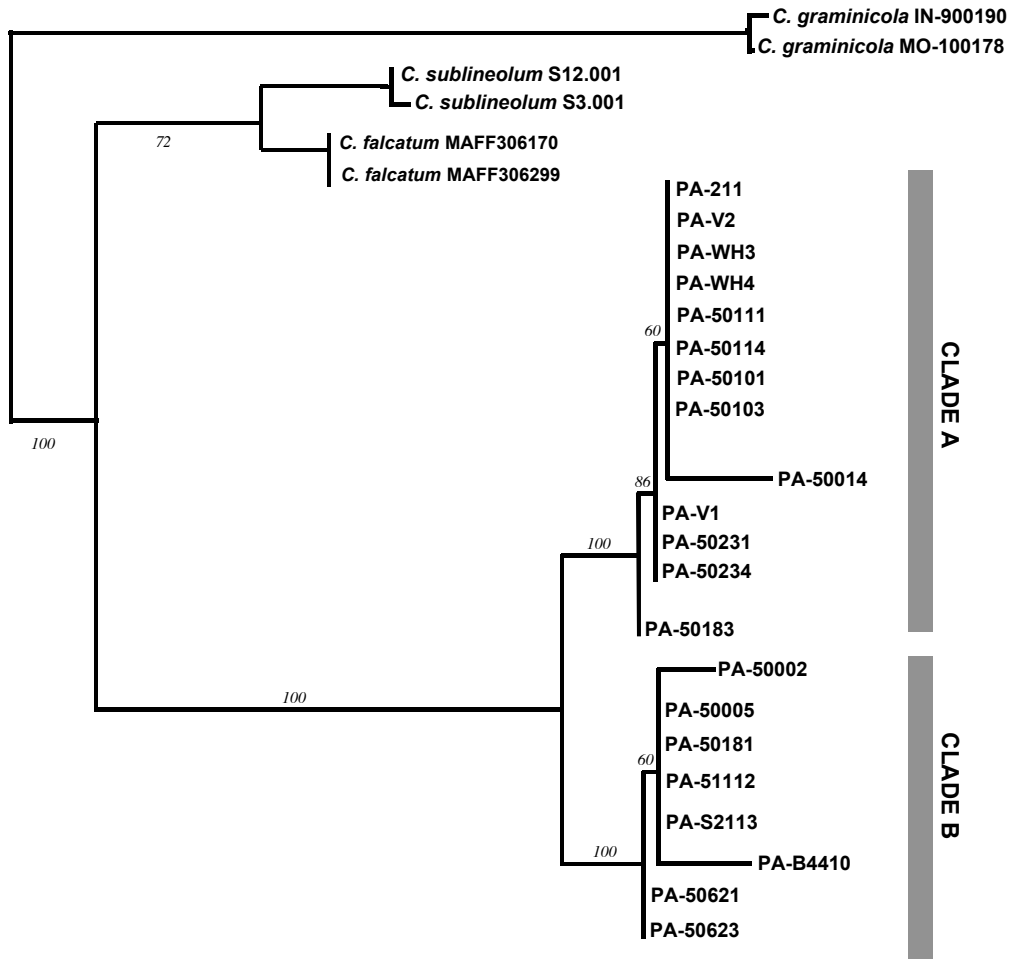


Figure 3.2
Multilocus tree estimated through Bayesian phylogenetic analysis of three protein coding genes supporting the division of the *C. cereale* isolates into two main lineages, clades A and B (*-ln* likelihood=3430.89).

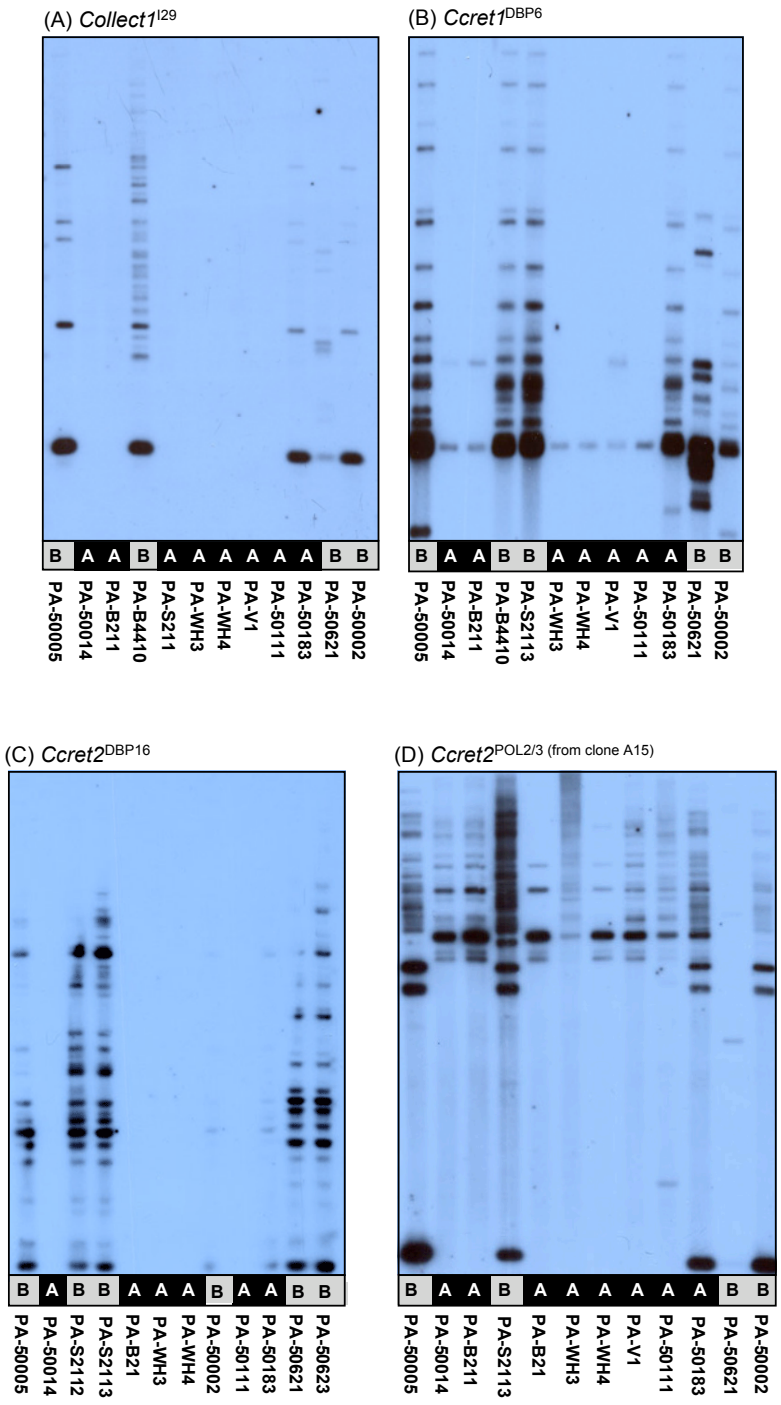


Figure 3.3

Southern blot hybridizations of *Hind*III digested genomic DNA from a representative sample of *C. cereale* clade A and B isolates using four transposon sequences as the probe. (A) *Collect1*^{I29} DNA transposon; (B) *Ccret1*^{DBP6} retrotransposon; (C) *Ccret2*^{DBP16}retrotransposon; (D) *Ccret2*$^{POL2/3}$ (from clone A15) retrotransposon.

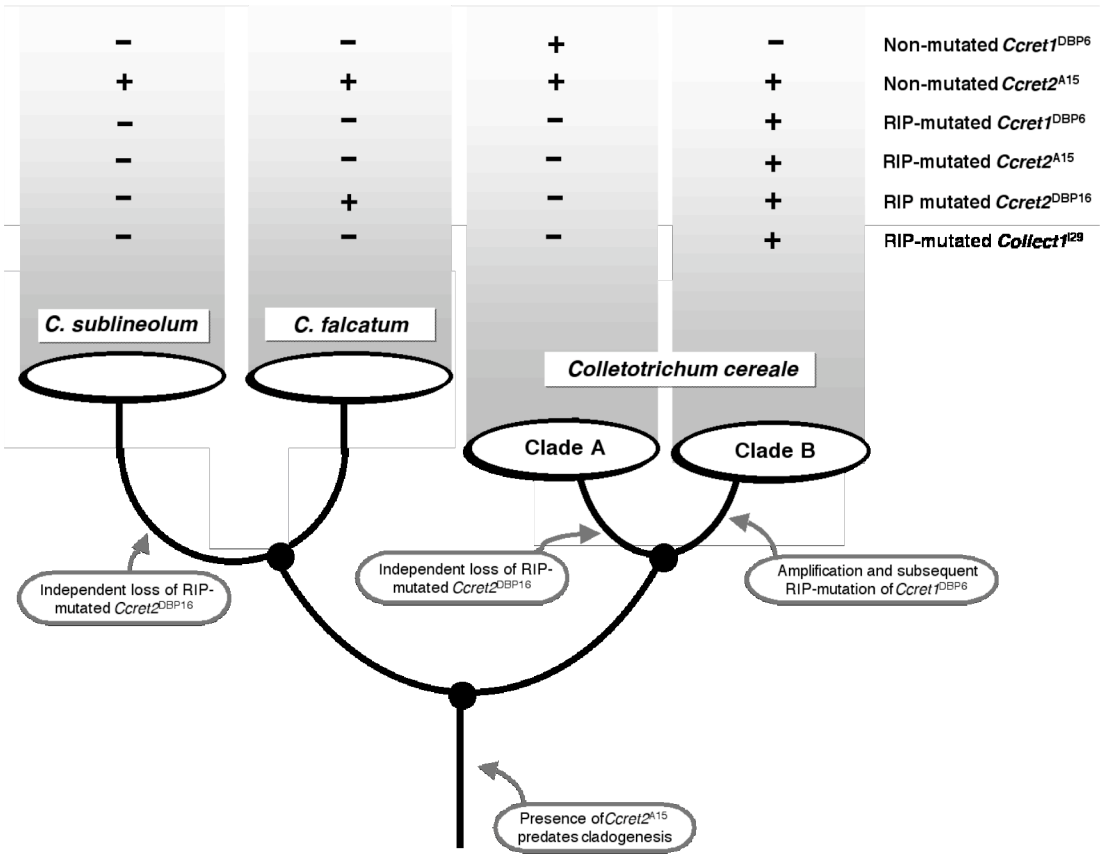


Figure 3.4
A schematic tree showing the presence or absence of the transposons evaluated in this study.

Chapter 4: Ecological specialization drives the evolution of the pathogenic fungus *Colletotrichum cereale* and allied taxa across diverse grass communities

ABSTRACT

Emerging infectious diseases are problematic in both cultivated and natural plant communities, threatening food production, wildlife habitats and biodiversity conservation. Over the past decade, the emergence of anthracnose disease has newly challenged the health of turfgrasses on North American golf courses, resulting in considerable economic loss. The fungus responsible for the outbreaks, *Colletotrichum cereale*, has also been identified from numerous natural grasses and cereal crops, although disease symptoms are generally absent. Here we utilize phylogenetic and population genetic analyses to determine the role of ecosystem in the advancement of turfgrass anthracnose and assess whether natural grass and/or cereal inhabitants are implicated in the epidemics. Using a four gene nucleotide dataset to diagnose phylogenetic species and population boundaries, we find that the graminicolous *Colletotrichum* diverged from a common ancestor into two distinct lineages correspondent with host physiology (C3 or C4). In the C4 lineage, which includes the important cereal pathogens *C. graminicola*, *C. sublineolum*, *C. falcatum*, *C. eleusines*, *C. caudatum* and several novel species, host specialization predominates, with host-associated lineages corresponding to isolated phylogenetic species. In contrast, although the C3 lineage -- *C. cereale* -- is comprised of one wide host-range species, it is divided into ten highly specialized populations corresponding to

ecosystem and/or host plant, all derived from a generalist founder population. Extreme differentiation between the specialized *C. cereale* populations suggests that asymptomatic non-turfgrass hosts are unlikely reservoirs of infectious disease propagules, but gene flow between the generalist founder population and the specialized genotypes provides an indirect mechanism for genetic exchange between otherwise isolated populations and ecosystems. These findings demonstrate that while disease occurrence and spread is currently localized to the turfgrass environment, introgression between *C. cereale* ecotypes can lead to the expansion of anthracnose disease into new ecosystems.

4.1 Introduction

The emergence of infectious plant disease epidemics over the course of the last century has substantially impacted global biodiversity, shaping the community structure of cultivated landscapes and natural ecosystems alike (Holmes, 1996). Disease-inducing fungi, bacteria and viruses exert considerable influence over the fitness and survival not only of the infected plant but also that of animals, including insects and birds, or any other organism reliant upon the host species for food or shelter. During the early 20th century, for example, introduction of the chestnut blight fungus, *Cryphonectria parasitica*, from Asia into the United States almost drove the American chestnut (*Castenea dentata*) to extinction in North American forests (Anagnostakis, 1988; Milgroom, 1995; Milgroom *et al.*, 1996), dramatically restructuring the forest ecosystem and leading to the virtual extinction of several phytophagous insect species (Opler, 1979). While the emergence of pathogens may be associated with various inciting factors, novel plant disease epidemics

are primarily driven by the anthropogenic introduction of parasites (Anderson *et al.*, 2004). Movement of novel organisms into naive environments provides the opportunity for rapid adaptive evolution, often due to novel or episodic selective pressures (Brasier, 1995), increased fitness as a result of niche differentiation (Ennos, McConnell, 1995), or by means of introgressive hybridization imparting elevated pathogenicity or enhanced fitness (Abdelai *et al.*, 1999). But in this era of unparalleled environmental change, human-driven modification of terrestrial ecosystems and severe weather events are increasingly important contributors to disease outbreaks, acting either individually or in concert (Anderson *et al.*, 2004).

Beginning in the 1990s, anthracnose disease caused by the haploid fungus *Colletotrichum cereale* Manns (*sensu lato* Crouch, Clarke and Hillman [Crouch *et al.*, 2006]) emerged as one of the most destructive diseases of turf-type grasses maintained as golf course greens in the United States and Canada. Since that time, the incidence, severity and geographic range of turfgrass anthracnose greatly expanded (Smiley *et al.*, 2005). Although anthracnose disease ascribed to *C. cereale* was first demonstrated in association with cereals and grasses of the Pooideae subfamily in 1908 (Selby, Manns, 1909), the fungus was deemed a trivial pathogen throughout the next seven decades. Reports of anthracnose disease outbreaks were recorded during the course of the 20th century, with the disease making transient appearances in turfgrass, cereal rye, wheat, oats, and orchardgrass in the United States, England, Brazil and South Africa, with varying degrees of severity (Baxter *et al.*, 1983; Kemp *et al.*, 1991; Minussi, Kimati, 1979; Selby, Manns, 1909; Smith, 1954). During the 1970s and 1980s, turfgrass pathologists debated whether anthracnose disease symptoms on turf were even induced by *C. cereale* (then referred to as *C. graminicola*), speculating that the presence of the fungus in association with unhealthy

plants was likely coincidental and suggesting instead that declines in plant health were due to either abiotic factors or the presence of a second fungus, *Helminthosporium* spp. (Smiley *et al.*, 2005). Further adding to the confusion and uncertainty surrounding this organism was the fact that between 1914 through 2006, because of extreme morphological similarity, *C. cereale* was considered conspecific with *C. graminicola*, an important pathogen of corn (Wilson, 1914). Notwithstanding these modest origins, *C. cereale* is now confirmed as the causal agent of anthracnose outbreaks on golf courses throughout North America, presenting one of the most significant challenges to the health of *Poa annua* (annual bluegrass) and, increasingly, *Agrostis stolonifera* (creeping bentgrass) turfgrasses. Unfortunately, because this organism and the disease it causes are largely unknown, control of anthracnose on turfgrass is heavily reliant upon substantial inputs of costly fungicidal chemicals, resulting in widespread fungicide resistance across *C. cereale* populations (Crouch *et al.*, 2005; Wong, Midland, 2007; Wong *et al.*, 2007).

Why has *C. cereale* traded its long history of relative anonymity to assume the role of a destructive plant pathogen? Outside of anecdotal comments or predictions made through the study of other species of *Colletotrichum*, almost nothing is known about the biology of *C. cereale*, including the mechanisms responsible for infection of the host plant, overwintering strategies or the generation and dissemination of reproductive propagules. Direct observation of the fungus *in vitro* suggest that mitotic spores are the primary, and perhaps only, means of propagation for the fungus (Browning *et al.*, 1999), J.A. Crouch, B.I. Hillman, unpublished data). Rarely, laboratory matings produce malformed protoperithecia containing misshapen asci, but these spore sacs have never been found to contain the meiotic ascospores (J.A. Crouch, B.I. Hillman, unpublished data). Molecular analysis, however, revealed the signature of recombination acting upon at least one *C.*

cereale population, suggesting that some process facilitates the exchange of genetic material between individuals (Crouch *et al.*, 2006). Given the recent identification of transposons altered by repeat-induced point mutation (Crouch *et al.*, 2007), a fungal genome defense mechanism that occurs only during meiosis, it is likely that the observed recombination is sexually derived, despite the physical absence of a viable sexual morph.

Population structure and genetic diversity are similarly obscure for this organism, although for North American isolates of *C. cereale* from turfgrass hosts, two morphologically cryptic lineages that possess overlapping host and geographic ranges are well documented (Crouch *et al.*, 2005; Crouch *et al.*, 2006; Crouch *et al.*, 2008a; Crouch *et al.*, 2008b). Importantly, molecular phylogenetic analysis demonstrated that *C. cereale* is not solely limited to the lifestyle of a turfgrass pathogen, with non-pathogenic isolates of the fungus detected from asymptomatic cool-season (C3) prairie grasses (*Dactylis glomerata*, *Elymus virginicanus*, *E. canadensis* and *Bromus inermis*) and wheat crops (*Triticum aestivum*) (Crouch *et al.*, 2006). As a result, a continuous distribution of suitable host plants across the continent may provide a sympatric landscape for *C. cereale*, with pathogen and non-pathogen populations sharing an overlapping range that spans most of North America. But are all grass communities equally suitable *C. cereale* habitats? This is a question with enormous implications for the initiation and expansion of turfgrass anthracnose disease, and may have influenced the emergence of this fungus as a pathogen during the 1990s. If all *C. cereale* isolates are uniformly fit across all grass community types, irrespective of genotype or pathotype, then natural stands of asymptomatic grasses could potentially serve as reservoirs of infectious disease particles, continuously fueling the disease epidemics in cultivated turfgrass systems. Given the tremendous disparity in the environments that the fungus might inhabit, we hypothesized that specialization to the

cultivated turfgrass ecosystem might distinguish populations of the fungus responsible for turfgrass anthracnose epidemics from their non-pathogenic counterparts. In stands of cultivated monocultures of perennial turfgrasses maintained as golf course greens, host plant populations are extremely dense and genotypically regular, and are maintained under a stress-inducing regimen of low fertility, low cutting height, water deficiency and mechanical injury while being simultaneously subjected to pesticides and plant growth regulating chemicals. Cultivated annual agroecosystems like those found in wheat fields, while also relatively genetically uniform, are much less densely planted than turf and are subjected to fewer chemical and mechanical inputs. In contrast to the man-made environments, natural grass ecosystems are complex and genetically diverse, composed of both annual and perennial grasses and dicotyledonous plants, with the distribution of host plants sometimes patchy, and an absence of chemical inputs or mechanical disturbances.

In this study we utilize multi-locus nuclear sequence data to explore the hypothesis that specialization of *C. cereale* to the unique turfgrass environment might limit the movement of the fungus from cereal crops and prairie grasses. Because species boundaries are presently poorly defined in the genus *Colletotrichum*, we considered both phylogenetic and population genetic patterns of habitat specialization that may have served to differentiate *C. cereale* populations not only from one another, but also from several sympatrically distributed, grass-associated *Colletotrichum* species that are presumed to be *C. cereale's* closest relatives. Two broad questions were specifically addressed in the course of this research; each one serves to expand the underlying hypothesis. First, to what extent are species and population boundaries in the graminicolous *Colletotrichum* determined by the ecosystem which the fungus inhabits? Secondly, are there associations between genotypes and the origins of populations that might account for the recent

emergence of *C. cereale* as a pathogen of North American turfgrass?

4.2 Materials and methods

4.2.1 Sample collection

SI tables 1 and 2 list the 208 samples of *Colletotrichum* isolated from 38 species of grasses evaluated in this study. Two samples of *C. acutatum* were included as outgroup taxa to root the phylogenetic trees (Crouch *et al.*, 2006). *Colletotrichum* samples were obtained by performing isolations of the fungus from infested grass tissue as previously described (Crouch *et al.*, 2005) or from culture collections. For this research, *C. cereale* samples were broadly cataloged as (1) pathogenic to their host, based upon evidence of anthracnose disease in the plant from which they were isolated; (2) non-pathogenic to their host, based upon the complete absence of overt disease symptoms in the host plant from which they were isolated; or (3) in the case of samples obtained from culture collections where lifestyle data were not available, of undefined pathogenicity to the host from which they were isolated (SI Table 1). No pre-epidemic North American turfgrass-derived strains of *C. cereale* could be located for this study. Fungal samples were maintained in culture on potato dextrose agar (PDA; Fisher Scientific, Hampton, NH) or stored dessicated at -70^{o}C adhered to silica gel particles (Davis, 2000). DNA was purified from single-spore purified dehydrated fungal tissue grown on cellophane disks plated over PDA using a standard phenol:chloroform extraction protocol (Sambrook *et al.*, 1989).

4.2.2 Molecular data

Previous work showed that sequence analysis of 1229-bp from three loci was sufficient to address the taxonomic distinction between *C. cereale* and the morphologically similar grass pathogens *C. graminicola* and *C. sublineolum*, but could not distinguish the inter-relatedness of these taxa, nor were those data capable of differentiating between *C. cereale* isolates beyond the identification of two major lineages (Crouch *et al.*, 2006). To obtain data for fine-scale resolution of both phylogenetic and population-level relationships, four nuclear genes from three genomic regions were sampled to generate 3031-bp of sequence. Analysis of two regions adjacent to the *C. cereale* mating type locus (*Mat1*) showed a high level of polymorphism among isolates of *C. cereale* and variability between species which made them especially suitable as markers for studying evolutionary patterns at the boundary between populations and species. Although in many fungi *Mat1* cannot be employed as a phylogenetic marker due to the presence of highly dissimilar alleles at the locus (*Mat1-1* or *Mat1-2*), all *Colletotrichum* sampled to date possess the *Mat1-2* sequence, making it a suitable marker for this group (*e.g.* Du *et al.*, 2005; Rodríguez-Guerra *et al.* 2005; Crouch *et al.*, 2006). The single copy"W1" sequence spans ~850-bp of the 5' end of *Apn1*, the apurinic DNA lyase adjacent to the *Mat1* sequence (J.A. Crouch, L.J. Vaillancourt, B.I. Hillman; unpublished data). The single copy "M72" sequence is ~1,400-bp and spans the 3' end of *APN1*, the intergenic region between *Apn1* and *Mat1*, and the 3' end of *Mat1* (J.A. Crouch, L.J. Vaillancourt, B.I. Hillman; unpublished data). Primers used to amplify the W1 and M72 sequences were: W1F: 5' ATGGAGCACAAAAACGAACA 3'; W1R: 5'GCGGAGCAGAGGATGTAGTC 3'; and M72F: 5' ACGGCAAACGGCTCAGGGAGTG 3'; M72R:

AATGCCGAGTCCCACGAGGTTCG 3'. The W1 and M72 products were PCR amplified from genomic DNA using an initial denaturation cycle of 95°C for 5 minutes, followed by 40 cycles of a 95°C melt step for 60 seconds; an annealing step between 55-48°C for 60 seconds (temperature dependant upon individual sample requirementss), an extension at 72°C for 1 or 2 minutes (W1 or M72, respectively), followed by a final extension of 20 minutes. PCR amplification of the ~450-bp internal transcribed spacer (ITS) region of the ribosomal DNA using the ITS4/5 primer pair and the ~550-bp single copy manganese superoxide dismutase (*Sod2*) gene using primers SOD625F/R or SOD507F/R (*SOD2*) were performed as described (Crouch *et al.*, 2006; White *et al.*, 1990). PCR amplicons were gel purified using the GeneCleanIII kit (Qbiogene, Irvine, CA) and used as templates for sequence reactions performed with the Big Dye® v3.01 chemistry (Applied Biosystems, Foster City, CA) at 1/12 the manufacturer's recommended reaction on an ABI 3100 automated sequencer for 99 cycles; amplicons were sequenced in both directions. Sequences were edited using the LaserGene DNAStar software suite (Madison, WI). Multiple sequence alignments were constructed using the program M-Coffee (Wallace *et al.*, 2006), followed by manual refinements according to predicted amino acid sequences. Gaps were re-coded as multistate characters; in regions where the alignment was ambiguous, the questionable sections were excluded from the dataset.

4.2.3 Phylogenetic analyses

Phylogenetic trees were constructed from separate and combined analysis of the four gene regions using maximum likelihood (ML) and Bayesian methods; maximum

parsimony (MP) analyses were also performed from the individual gene sequences for comparative purposes. ML searches were executed through the metapopulation genetic algorithm implemented in MetaPIGA v1.0.2 (Lemmon, Milinkovitch, 2002), with each search replicated 100 times. MP analyses were executed in PAUP* using heuristic searches, 10 random addition sequence replicates, and TBR branch swapping. Heuristic parsimony searches were performed using 100 random taxon addition sequence replicates and tree-bisection reconnection (TBR) branch swapping; multistate characters were treated as polymorphisms; and support for individual clades was evaluated using 500 bootstrap pseudoreplicates consisting of 10 random addition sequence replicates. The best fit model of DNA sequence evolution, as identified using the AIC in ModelTest v.3.7 (Posada, Crandall, 1998) was incorporated into the likelihood and Bayesian analyses (ITS: SYM+G, base freq.=equal, Nst=6, Rmat=1.56/1.42/1.70/3.13/3.33/1.0, I=0, α=1.16; M72: TIM+I+G, base freq.=0.24/0.29/0.26/0.21, Nst=6, Rmat=1.0/2.23/0.70/0.70/3.18/1.0, I=0.14, *Sod2*: TVM+I+G, base freq.=0.24/0.31/0.28/0.72, Nst=6, Rmat=0.73/3.56/1.58/0.59/3.56/1.0, I=0.36, α=0.99; *Sod2*: TVM+I+G, base freq.=0.24/0.31/0.28/0.72, Nst=6, Rmat=0.73/3.56/1.58/0.59/3.56/1.0, I=0.36, α=0.99; W1: TIM+G, base freq.=0./24/0.29/0.25/0.23 Nst=6 Rmat=1.0/3.84/0.57/0.57/5.29/1.0, α=0.63, Pinvar=0). Bayesian posterior probabilities supporting the ML tree topologies were generated using Mr. Bayes v3.1 (Huelsenbeck, Ronquist, 2003) by performing two simultaneous runs of one cold and three incrementally heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC) for 1,000,000 generations and sampling trees every 500 generations. Trees drawn from the posterior probability distribution were identified using Tracer v1.3 (Rambaut, Drummond, 2006), then imported into PAUP* v.4.0b10

(Swofford, 2000) and used to construct a 50% majority-rule consensus tree from which posterior probability-based support values were derived. Species were diagnosed within a phylogenetic context, using a modified multilocus genealogical concordance approach as described by Taylor and colleagues (Dettman *et al.*, 2003; Taylor *et al.*, 2000). To produce a phylogenetic species, an ancestral lineage splits into two evolutionarily independent descendant lineages, thus, a phylogenetic species must meet four criteria: (1) monophyly; (2) strong support values from both bootstrap and Bayesian posterior probabilities in the combined dataset analysis; (3) genealogical concordance; i.e. there could be no conflict from one gene tree to the next, or between an individual gene tree and the tree inferred from the combined dataset; and (4) no evidence of recombination with individuals outside the group.

4.2.4 Population analyses

For population analyses, the combined nucleotide dataset was pruned of all non-variable characters using Collapse v1.2 (http://darwin.uvigo.es/software/collapse.html), resulting in 96 unique *C. cereale* haplotypes and 197 characters. Population subdivision was identified using (1) visualization of well-supported phylogenetic lineages (above) and (2) Bayesian assessment of populations using the Bayesian Monte Carlo Markov chain-based clustering program STRUCTURE 2.1 (Falush *et al.*, 2003; Pritchard *et al.*, 2000). STRUCTURE analyses were performed for 20 runs of 1,000,000 repetitions each, for K=1 through 20 (where K=the maximum number of populations), with the first 20,000 repetitions discarded as burn-in. These analyses were run using the admixture model and

correlated allele frequencies between populations; the degree of a admixture was empirically derived from the data, and the distribution of allelic frequencies was set to 1 (Falush *et al.*, 2003). To visualize the intraspecific evolutionary relationships between individuals and populations, a median-joining network was constructed for the *C. cereale* haplotype dataset using the software NETWORK (www.flexus-engineering.com/netwinfo.htm). Partitioning of variance within and between populations was evaluated using AMOVA, and pairwise measures of gene flow/migration (N_em) and differentiation between populations were estimated using the F-statistic analog PhiPT (Excoffier *et al.*, 1992) as implemented in GenAlEx (Peakall, Smouse, 2006). Recombination was estimated through calculation of the Phi recombination estimator statistic (ϕ_w) (Bruen *et al.*, 2006) performed using SplitsTree v4.8 (Huson, 1998).

4.3 Results

4.3.1 Species boundaries and the evolution of grass-associated *Colletotrichum*

We first established the evolutionary origins, relationships to other grass-inhabiting *Colletotrichum* species and host range limits of *C. cereale* by constructing a multilocus gene genealogy from 208 samples of *Colletotrichum* isolated from grasses. Representatives from all six known species of *Colletotrichum* associated with grass hosts (*C. cereale*, *C. graminicola*, *C. falcatum*, *C. sublineolum*, *C. caudatum* and *C. eleusines*) were included in the analysis, along with 50 samples without certain species assignment (SI Tables 1 and

2). In total, 65 *Colletotrichum* samples from 14 species of warm-season (C4) grasses and 143 samples from 24 species of C3 grasses from the grass subfamily Pooideae, collected from 12 countries from 1934 through 2006, were evaluated (Fig. 1).

The four gene, three locus nucleotide sequence dataset (ITS, *Sod2*, *Apn1* and *Mat1*) employed in this work provided substantial resolution of species and populations within the group of grass-associated *Colletotrichum* (Fig. 2). In particular, the inclusion of 2088-bp of sequence data from the *Apn1* and *Mat1* genes and the connecting intergenic region provided considerable phylogenetic signal (SI Figs. 1-4). Using parsimony statistics from the ingroup taxa as a comparative benchmark, the 1290-bp M72 (*Apn1/Mat1*) and 798-bp W1 (*Apn1*) sequences were 58% and 72% informative, while the 536-bp *Sod2* and 407-bp ITS sequences were 40% and 50% informative, respectively. Phylogenetic reconstructions using the combined 3031-bp dataset yielded a strongly supported pattern of divergent evolution between the *Colletotrichum* sampled from C4 and C3 grasses (Fig. 2), with the physiology of the grass host entirely connected within these two primary fungal lineages. The tree topology shown in Figure 2 depicted a slightly earlier divergence of populations for the C3-associated lineage relative to the divergence of species in the C4 lineage, although the distance observed between the divergences of populations and species within these two groups suggested that a simultaneous radiation was an equally likely sequence of events. The C4 grass clade was comprised exclusively of C4 grass-associated pathogens, with the host range for these samples drawn from grasses of the PACCAD grass clade, primarily within the subfamilies Panicoideae and Paniceae (Fig 2). Despite the morphological similarity of the *Colletotrichum* represented in the C4 lineage (Sutton, 1980; Sutton, 1992), these taxa showed considerable diversity using molecular characters. While only five species have been previously described in association with C4

grasses based upon morphological characters (*C. graminicola*, *C. falcatum*, *C. sublineolum*, *C. caudatum*, and *C. eleusines*) and only four species names are in common usage (Sutton 1982), the use of molecular phylogenetics to delineate species boundaries in the present study not only confirmed these morphospecies, but also identified eight additional distinct lineages definable as phylogenetic species (Fig. 2). Comparison of *Colletotrichum* species boundaries against host plant derivations showed the two factors were interconnected in the C4 lineage, with two phylogenetic species extracted from samples of *Saccharum officinarum* (sugarcane; *C. falcatum* and an undescribed species from sugarcane), two phylogenetic species from *Paspalum dilatatum* (dallisgrass; both undescribed phylogenetic species), and one phylogenetic species each from *Sorghum bicolor* and *S. halapense* (sorghum; *C. sublineolum*), *Zea mays* (corn; *C. graminicola*), *Eleusine indica* (goosegrass; *C. eleusines*), *Axonopus affinis* (carpetgrass; undescribed phylogenetic species), *Digitaria* sp. (crabgrass; undescribed phylogenetic), *Echinochloa esculenta* (Japanese millet; undescribed phylogenetic species), *Miscanthus sinensis* (maidengrass; undescribed phylogenetic species), and *P. notatum* (bahiagrass; undescribed phylogenetic species). Only *C. caudatum*, which, in addition to its distinctiveness on the molecular level, also possesses a unique conidial appendage, was comprised of isolates from multiple hosts: *Bothriochloa bladhii* (Caucasian bluestem), *Imperata cylindrica* (cogongrass), and *Zoysia tenuifolia* (Mascarene grass). Overall, a history of cladogenesis driven by host substrate is inferred from phylogenetic evidence in the C4 grass-associated lineage, suggesting a hypothesis of coevolution of this group of fungi with its hosts.

While cladogenesis within the C4-associated group was clearly depicted in both the likelihood and parsimony reconstructions, only three supra-specific clades were consistently well-supported, as the C4 group basal clade organization changed from one

gene genealogy to the next (SI Fig. 1-4). Because of this larger-scale conflict between genealogies, a good candidate as a founder for the C4 group could not be identified. Even comparisons between the W1 and M72 sequences, where the data were derived from two adjacent genes in the fungal mating type gene cluster and separated by less than 500-bp, there was conflict in the most basal branching organization, possibly emanating from incomplete lineage sorting between closely related taxa. Nevertheless, three supra-specific groupings of C4 species were consistent between the gene genealogies and the combined tree topology (Fig. 2; SI Figs. 1-4). C4 group 1 was comprised of three novel phylogenetic species derived from the Panicaceae grasses *Digitaria* sp., *Ec. esculenta* and *Ax. affinis*. C4 group 2 was comprised of *C. graminicola* from corn and three novel phylogenetic species from *Paspalum* sharing a common ancestry. In a third supra-specific group, *C. sublineolum* from sorghum and *C. falcatum* from sugarcane evolved from a common ancestor. Based upon host-derivation criteria, the causal agent of red rot disease in sugarcane, *C. falcatum*, was recovered as a polyphyletic taxon, with strains of this fungus exhibiting considerable sequence diversity and spanning two phylogenetic species.

4.3.2 Differentiation among *C. cereale* populations

The host origination and geographical distribution of *Colletotrichum* sampled from C3 Pooideae grasses in this analysis was very diverse (Fig. 1, SI Table 1); however, on the molecular level the C3-associated lineage showed considerably less variability than that found in the C4 (SI Fig. 5). Unlike the C4 lineage, which consisted of fourteen distinct phylogenetic species linked to host derivation, the C3 lineage was comprised of only a single species, *C. cereale*, associated with multiple hosts. Although eleven well-supported *C.*

cereale subgroups were revealed in the multilocus phylogeny (SI Fig. 5), these groups could not meet at least one of the criteria for the definition of phylogenetic species -- the absence of gene flow. Analysis of the *C. cereale* group using the ϕ_w statistic showed a statistically significant level of recombination in all datasets (ITS: p=2.32 e-4; M72: p=0.0; *Sod2*: p=2.67 e-5; W1: p=4.27 e-13; combined: p=0.0). Evidence of recombination from the current dataset and the results of previous work (Crouch *et al.*, 2006; Crouch *et al.* 2008a) supported the interpretation that the C3-associated lineage, i.e., *C. cereale*, comprised a single phylogenetic species.

Because recombination placed the *C. cereale* isolates outside of a strictly phylogenetic framework, intraspecific variation and population composition was evaluated using median-joining network analysis of the sample reduced to 96 distinct haplotypes from 143 *C. cereale* isolates (SI Table 1) and 197 variable characters from the 3031-bp nucleotide dataset. Eleven major *C. cereale* populations, consistent with the major phylogenetic-derived clusters, were detected (Fig. 3). The network topology illustrated an interwoven web of relationships consistent with recombination linking both individuals and populations, although clonal haplotypes were observed in a few groups. In the present study, C. *cereale* groups A2, A3, A6, A7, A9 and A10 correspond to isolates previously described as part of the larger *C. cereale* clade A; group B corresponds to the previously described *C. cereale* clade B (Crouch *et al.*, 2006); and members of groups A1, A4, A5, and A8 have not been previously sampled but fall within the circumscription of *C. cereale* clade A.

C. cereale was identified in association with 13 additional cool-season grass species, including the economically important cereal crops *Avena sativa* (oats), *Hordeum pusillum* (barley), and several prairie grasses (Fig. 3; SI Fig. 5; SI Table 1), bringing the known host

range of *C. cereale* to 24 grasses of the Pooideae. Despite this broad host range, there was a remarkable level of population subdivision that corresponded with particular hosts or ecosystem types (Fig. 3). Four distinct populations of *C. cereale* pathogenic to North American turfgrasses (A7, A9, A10 and B) were identified by each of the analytical methods applied to the multilocus dataset, with three of the populations (A7, A9 and A10) limited entirely to turfgrass hosts. In the three turfgrass-only populations, a clear association with a single host plant was evident: 83% of A7 strains were from annual bluegrass; 91% of A9 strains were from creeping bentgrass and 87% of A10 turfgrass strains were from annual bluegrass. In contrast, within the highly divergent group B, which also included a number of non-turfgrass derived members among its ranks (30%), turfgrass isolates were almost equally divided between annual bluegrass and creeping bentgrass. Similarly, non-turfgrass haplotypes from prairie grasses and cereal crops were exclusive to their own populations (A1, A2, A3, A4, A5, A6 and A8); but prairie isolates (A2, A3 and A5) did not dissociate from cereal isolates. Wheat isolates were, however, distributed in separate populations (A2, A3) from oat isolates (A1, A5). Thus, with the exception of the broadly distributed group B, populations were derived from either one of two primary environmental types -- the turfgrass pathogens, and the non-turfgrass isolates, which were either non-pathogenic or of undefined pathogenicity. In the tree topologies reconstructed from the combined dataset and the M72 sequences, group B was identified as *C. cereale's* founder population (SI Figs. 2, 5). The *Sod2*, W1 and ITS phylogenies did not possess enough signal to either confirm or reject the ancestral position of the B population, but did authenticate the distinctness of this group from clade A populations (SI Figs. 1, 3, 4). Previous reports of the clade A turfgrass haplotype 5 (H5) as a potential ancestral haplotype from which all *C. cereale* isolates might have radiated

(Crouch *et al.*, 2006) were not supported using the larger, more diverse sample from multiple ecosystems and the extremely variable dataset employed in the present study, where clade A H5 isolates (represented by haplotypes 53, 62, 63, 67, 68, 69 and 70 in this study) were dispersed between groups A9 and A10, depending upon whether the isolate in question was derived from annual bluegrass or creeping bentgrass hosts (SI Table 1, Fig. 3).

As evidenced by the reticulating network topology, high levels of genetic diversity and the ϕ_w statistic, the evolution of *C. cereale* has been strongly impacted by recombination. But, consistent with *C. cereale*'s prolific production of asexual spores, there were ample signs of clonal reproduction. In particular, a single haplotype dominated turfgrass group A7 and, to a lesser degree, A9 (Fig. 3), showing that in the turfgrass ecosystem, the fungus is reliant upon asexual reproduction through the generation of mitotically generated conidia to stimulate disease.

None of the *C. cereale* populations showed any association with geographic origin, although sampling was not specifically designed to test such an association. In the four turfgrass pathogen populations of North American origin, each was comprised of a trans-continental assemblage of *C. cereale* isolates (SI Table 1). The absence of geographic correlation was even more marked in the non-turfgrass cohort, which included a sample of both North American and international isolates (Japan, New Zealand, Germany). Like their turf counterparts, the non-turfgrass populations clustered according to ecosystem rather than geographical origin, with non-turf North American isolates more closely related to non-turf international isolates than to the North American *C. cereale* isolates responsible for the turfgrass anthracnose epidemic (Fig. 3; SI Table 1). Association based upon host was robust over the course of 70 years, as the non-turfgrass groups were

comprised of isolates spanning the period between 1934-2006 (SI Table 1). In combination with international distributions (USA, Canada, Japan, Germany, Netherlands and New Zealand), these data demonstrated that *C. cereale* genotypes A1, A2, A5, A8 and B have been globally dispersed for at least six decades.

To investigate the level of intraspecific divergence between populations, an AMOVA was performed as shown in Table 1. With the groups defined according to genetic populations, the AMOVA showed that among-population variability accounted for the majority of the variance in the dataset (65%), particularly in groups A5 and B, where the host composition was diverse in comparison to other *C. cereale* populations. Notably high levels of migration/gene flow suggested that group A8 is likely to be part of group B, and groups A1, A2 and A3 might be accurately collapsed into a single larger population. Overall and pairwise fixation index values (PhiPT) were significant in all but four pairwise combinations (93%) and revealed the majority of populations as exceptionally differentiated from one another (Table 2). Extreme isolation was observed between the two major specialized annual bluegrass turfgrass pathogen groups (A7 and A10), despite -- or perhaps because of -- their shared habitat (Table 3). The differentiation of annual bluegrass turfgrass pathogen group A10 was particularly noteworthy, in that this population was almost completely isolated from all other populations except for turfgrass-containing groups A9 and B. Isolation of A7 from other *C. cereale* populations may be sustained by the clonal makeup of this group, where 22 isolates were encompassed by a single haplotype. In contrast, the creeping bentgrass turfgrass specialist A9 exhibited relatively high levels of gene flow between other populations, especially other turfgrass populations, but also with non-pathogenic ecotypes derived from prairie or cereal crops. While N_em values suggested that while gene flow

between the majority of populations is a somewhat rare occurrence, all *C. cereale* populations shared a connection with group B, exchanging, on average, approximately one migrant every 1.5 generations -- low level, yet enough to prevent differentiation (Slatkin, 1987).

4.4 Discussion and conclusions

4.4.1 The impact of environmental context and host specialization on the evolution of grass-inhabiting *Colletotrichum*

This study documents a striking hierarchical pattern of habitat specialization associated with the radiation of the grass-inhabiting *Colletotrichum.* Outgroup comparisons, combined with intimate monophyletic sister taxa relationships and a common falcate-shaped spore morphology (Sutton, 1980; Sutton, 1992), suggest that the involvement of *Colletotrichum* with grass hosts arose only once in this genus. Following the appearance of *Colletotrichum* on grasses, three levels of local adaptation at the host and/or ecosystem level occurred. At the first level is the historical supra-specific division that mirrors grass host physiology; where the species-rich C4-associated lineage diverges evolutionarily from the single-species C3-associated lineage. Second, adaptation relative to ecosystem/lifestyle (turfgrass vs. non-turfgrass, pathogen vs. non-pathogen) distinguishes all but one of the *C. cereale* populations. And at the third level, narrow specialization to a single grass plant species distinguishes eleven of the twelve C4-

associated *Colletotrichum* species and three of the four turfgrass pathogen populations of *C. cereale*. In light of these predominantly environmentally constrained patterns of genotypic isolation, we conclude that divergence in the graminicolous *Colletotrichum*, both past and present, is coupled with, and perhaps dependent upon, ecological context.

Given the observed phylogenetic patterns, can we conclude that the extant, highly specialized taxa comprising the grass-associated *Colletotrichum* group evolved from a generalist ancestor, capable of exploiting the resources of any available grass plant? Because of the common ancestry of the C3 and C4 groups, the possibility of a generalist forebearer is a reasonable hypothesis. But in truth, at present we are not capable of even adequately resolving whether the emergence of the C3 group predated the genesis of the C4 group, or if they radiated simultaneously. Furthermore, based on the present study, it is impossible to determine with confidence which lineages are the oldest within the C4-inhabiting group. Since recombination was not detected between any of the C4 species, the absence of basal resolution in all likelihood either emanates from a genuine lack of supra-specific structure, or through the confounding signal of shared ancestral polymorphisms retained between these closely related species. Therefore, while the prospect of a generalist ancestor seems likely, substantiating such an entity will require further inquiry into the evolutionary foundation of the graminicolous *Colletotrichum*.

4.4.2 The origin of *C. cereale* populations in diverse ecosystems

This research provides multiple lines of evidence that *C. cereale* group B represents the extant founder population for the species, from which all others have arisen, including

populations of this fungus responsible for the emergence of turfgrass anthracnose disease in North America. Given the recombinant nature of the datasets, the identification of group B as the basal population using phylogenetic outgroup comparisons without corroborating evidence would normally be suspect, however, two separate observations drawn from the population genetic analysis support this conclusion. First, as the most diverse population with respect to host range, lifestyle and genetic differentiation, group B is endorsed on a theoretical basis as the most ancestral population, predicated on the fact that older populations accumulate a greater number of polymorphisms relative to newly evolved populations when mutations occur at a non-accelerated rate (Dean, Ballard, 2004) -- an assumption which is not violated by this dataset (clocklike evolution could not be rejected for any of the four datasets using likelihood ratio tests; data not shown). Group B is further supported as the founding population by the fact that although the sympatric divergence of almost every *C. cereale* population was accompanied by extreme genetic differentiation, relatively high levels of gene flow unite group B individually with each separate population. Based upon this precise balance of diversification and connectivity, we submit that the generalist group B gave rise to the more specialized populations of *C. cereale*.

4.4.3 The influence of alternative *C. cereale* ecotypes on the initiation and maintenance of anthracnose disease in turfgrass and other ecosystems

The Pooideae grasses that serve as host to *C. cereale* comprise, both economically and ecologically, one of the most important plant families. The Pooideae count among

their ranks several vital cereal crops such as wheat, oats and barley; as well as natural grasses that are almost universally distributed throughout uncultivated environments, spanning prairies, meadows, woods, coastal areas, pastures, deserts, and mountain areas across the continent. Despite its outwardly diverse host range, the substantial division of *C. cereale* populations in accordance with host/ecosystem, in combination with extreme levels of genetic differentiation and a relatively large clonal contingent in turf populations, suggests that natural grasses and cereal crops are unlikely to serve as an alternate habitat for the three primary turfgrass disease-associated genotypes. Structuring and differentiation of the eleven *C. cereale* populations show that ecosystem-level specialization is a defining characteristic of turfgrass anthracnose disease in North America. The majority (87% of the present sample) of *C. cereale* isolates causing disease in turfgrass are drawn from three sympatric, highly specialized and genetically differentiated populations (A7, A9, A10). These three specialist turf populations are genotypically distinct from *C. cereale* isolated from non-turfgrass hosts. From this perspective, stands of asymptomatic non-turfgrass species probably do not serve as significant reservoirs of infectious spores that promote the maintenance and spread of turfgrass anthracnose epidemics. However, *C. cereale* generalist group B once again introduces an exception, in that genotypes within the circumscription of the B population are not limited to turfgrass environments. In concert with the considerable evidence of diversity, recombination and migration that shapes this population (Crouch *et al* 2006; Crouch *et al* 2008; this study), *C. cereale* group B isolates provide a unique opportunity for genetic exchange between differing populations drawn from different environmental backgrounds. Under favorable circumstances, recombination and introgression could provide the basis for a transition from benign grass inhabitants to invasive plant pathogens of ecosystems currently unaffected by

anthracnose disease. The data offered herein provide the crucial foundation needed to determine the extent of these processes, and, in combination with the development of a suite of polymorphic *C. cereale* microsatellite markers (J.A. Crouch, unpublished data) and expanded landscape level sampling, offer considerable opportunities for further study of the evolutionary diversification of *C. cereale* across varied grass communities.

4.5 References

Abdelai E, Brasier CM, Bernier L (1999) Localization of a pathogenicity gene in *Ophiostoma novo-ulmi* and evidence that it may be introgressed from *O. ulmi*. *Molecular Plant Microbe Interactactions* **12**, 6-15.

Anagnostakis SL (1988) *Cryphonectria parasitica*, cause of chestnut blight. In: *Advances in plant pathology* (ed. Sidhu GS), pp. 123-136. Academic Press, Ltd., London.

Anderson P, Cunningham A, Patel N, Morales, FJ, Epstein, PR, Daszak, P (2004) Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers. *Trends in Ecology and Evolution* **19**, 535-544.

Baxter AP, van der Westhuizen GCA, Eicker A (1983) Morphology and taxonomy of South African isolates of *Colletotrichum*. *South African Journal of Botany* **2**, 259-270.

Bergstrom, GC (2008) Anthracnose. In: *Compendium of Wheat Diseases*, in press, (Wiese MV, ed.). 3rd edn., APS Press, St. Paul.

Brasier CM (1995) Eposodic selection as a force in fungal microevolution with special reference to clonal speciation and hybrid introgression. *Canadian Journal of Microbiology* **73**, 1213-1221.

Browning M, Rowley LV, Zeng P, Chandlee JM, Jackson N (1999) Morphological, pathogenic, and genetic comparisons of *Colletotrichum graminicola* isolates from *Poaceae*. *Plant Disease* **83**, 286-292.

Bruen TC, Phillipe H, Bryant D (2006) A simple and robust statistical test to detect the presence of recombination. *Genetics* **172**, 2665-2681.

Crouch JA, Clarke BB, Hillman BI (2005) Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. *International Turfgrass Society Research Journal* **10**, 186-195.

Crouch JA, Clarke BB, Hillman BI (2006) Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. *Phytopathology* **96**, 46-60.

Crouch JA, Glasheen BM, Giunta MA, Clarke BB, Hillman BI (2008a) The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. *Fungal Genetics and Biology* **45**,190–206.

Crouch JA, Glasheen BM, Uddin W, Clarke BB, Hillman BI (2008b) Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles. *Crop Science*, in press.

Davis R (2000) *Neurospora: Contributions of a model organism*. Oxford University Press, New York.

Dean M, Ballard J (2004) Linking phylogenetics with population genetics to reconstruct the geographic origin of a species. *Molecular Phylogenetic and Evolution* **32**, 998-1009.

Dettman JR, Jacobson DJ, Taylor JW (2003) A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* **57**, 2703-2720.

Du, M., Schardl, C. L., Nuckles, E. M., and Vaillancourt, L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Colletotrichum* species complexes. *Mycologia* 97(3):641-58.

Ennos R, McConnell K (1995) Using genetic markers to investigate natural selection in fungal populations. *Canadian Journal of Botany* **73**, S302-310.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to mitochondrial DNA restriction data. *Genetics* **131**, 479-491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-1587.

Holmes JC (1996) Parasites as threats to biodiversity in shrinking ecosystems. *Biodiversity Conservervation* **5**, 975-983.

Huelsenbeck JP, Ronquist F (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Huson DH (1998) SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68-73.

Kemp GHJ, Pretorius ZA, Smith J (1991) Anthracnose of wheat in South Africa. *Phytophylactia* **23**, 177-179.

Lemmon A, Milinkovitch M (2002) The metapopulation genetic algorithm: An efficient solution for the problem of large phylogeny estimation. *Proceeding of the National Acadamy of Science USA* **99**, 10516-10521.

Milgroom MG (1995) Population biology of the chestnut blight fungus, *Cryphonectria parasitca*. *Canadian Journal of Botany* **73 (Suppl. 1)**, S311-S319.

Milgroom MG, Wang K, Zhou Y, Lipari SE, Kaneno S (1996) Intercontinental population structure of the chestnut blight fungus, *Cryphonectria parasitica*. *Mycologia* **88**, 179-190.

Minussi E, Kimati H (1979) Taxonomy of *Colletotrichum graminicola* (Ces.) Wils. (sensu Arx, 1957). *Review Centro Cienas Rurais* **9**, 171-187.

Opler PA (1979) *The American Chestnut Symposium*. Univ. of West Virginia Press, Morgantown, WV.

Peakall R, Smouse P (2006) Genalex 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6**, 288-295.

Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

Rambaut A, Drummond AJ (2006) Tracer: A program for analyzing results from Bayesian MCMC programs such as BEAST & MrBayes. Distributed by the authors, http://evolve.zoo.ox.ac.uk/software.html?id=tracer.

Rodríguez-Guerra R, Ramírez-Rueda MT, Cabral-Enciso M, García-Serrano M, Lira-Maldonado Z, Guevara-González RG, González-Chavira M, Simpson J (2005) Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. *Mycologia* **97**, 793-803.

Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory, New York.

Selby AD, Manns TF (1909) Studies in diseases of cereals and grasses. *Ohio Agrriculture Experimental Station Bulletin* **203**, 186-236.

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations, *Science* **236**, 787–792.

Smiley RW, Dernoeden PH, Clarke BB (2005) *Compendium of Turfgrass Diseases*, 3rd edn. APS Press, St. Paul.

Smith JD (1954) A disease of *Poa annua*. *Journal of the Sports Turf Research Institute*, 344-353.

Sutton BC (1980) *The coelomycetes: fungi imperfecti with pycnidia, acervuli, and stromata*. Commonwealth Mycological Institute, Kew, U.K.

Sutton BC (1992) The genus *Glomerella* and its anamorph *Colletotrichum*. In: *Colletotrichum: Biology, pathology and control* (eds. Bailey JA, Jeger MJ), pp. 1-26. CAB International, Wallingford, U.K.

Swofford DL (2000) PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Taylor JW, Jacobson DJ, Kroken S, *et al.* (2000) Phylogenetic species recognition and species concepts in fungi. *Fungal Genetics and Biology* **31**, 21-32.

Wallace I, O'Sullivan O, Higgins D, Notredame C (2006) M-Coffee: combining multiple sequence alignment methods with T-Coffee. *Nucleic Acids Research* **34**, 1692-1699.

White TJ, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: *PCR Protocols: A Guide to Methods and Applications* (eds. Innis MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315-322. Academic Press, San Diego.

Wilson GW (1914) The identity of the anthracnose of grasses in the United States. *Phytopathology* **4**, 106-112.

Wong F, Midland S (2007) Sensitivity distributions of California populations of *Colletotrichum cereale* to the DMI fungicides propiconazole, myclobutanil, tebuconazole and triadimefon. *Plant Disease* **91**, 1547-1555.

Wong F, Midland S, de la Cerda K (2007) Occurrence and distribution of QoI-resistant isolates of *Colletotrichum cereale* from annual bluegrass in California. *Plant Disease* **91**, 1536-1546.

Table 4.1. Summary AMOVA table, with population boundaries determined from the Structure analysis.

Source	d.f.	Sum of Squares	Mean Squares	Variance components	Percentage of variation
Among Populations	10	1109.69	110.97	13.80	65%
Within Populations	76	566.46	7.45	7.45	35%
Total	86	1676.15		21.26	100%

Table 4.2. Pairwise population PhiPT values, calculated as PhiPT = AP/(WP+AP)=AP/Total (AP=Estimated variance among populations, WP=estimated variance within populations). Average PhiPT for the sample=0.649 (p=0.001). PhiPT values are shown below the diagonal, probability values based on 9999 permutations are shown above the diagonal. Italicized values are not significant at p=0.05.

	A1 Prairie, barley & oats	**A2** Prairie & wheat	**A3** Prairie & wheat	**A4** Prairie	**A5** Prairie & oats	**A6** Prairie	**A7** Turf *Poa*	**A8** Prairie	**A9** Turf *Agrostis*	**A10** Turf *Poa*	**B** Mix
A1	--	0.00	0.03	0.04	0.00	0.01	0.05	0.05	0.00	0.00	0.00
A2	0.47	--	0.17	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A3	0.24	0.24	--	0.08	0.00	0.03	0.11	0.08	0.01	0.00	0.00
A4	0.78	0.90	0.75	--	0.01	0.07	0.33	0.34	0.01	0.01	0.00
A5	0.73	0.82	0.72	0.70	--	0.00	0.00	0.02	0.00	0.00	0.00
A6	0.87	0.93	0.86	0.96	0.63	--	0.07	0.08	0.00	0.00	0.00
A7	0.70	0.84	0.51	0.78	0.69	0.88	--	0.34	0.07	0.00	0.02
A8	0.62	0.86	0.56	0.59	0.58	0.73	0.51	--	0.05	0.01	0.01
A9	0.65	0.77	0.54	0.53	0.58	0.56	0.40	0.43	--	0.00	0.00
A10	0.90	0.93	0.89	0.91	0.79	0.91	0.83	0.85	0.53	--	0.00
B	0.49	0.63	0.43	0.43	0.53	0.46	0.41	0.30	0.44	0.61	--

Table 4.3. Pairwise population N_em values, shown below the diagonal, calculated as Nm = ((1/PhiPT)-1)*0.25. Average N_em for the sample=0.270. Italicized values are based upon non-significant PhiPT values (p=0.05).

	A1 Prairie, barley & oats	**A2** Prairie & wheat	**A3** Prairie & wheat	**A4** Prairie	**A5** Prairie & oats	**A6** Prairie	**A7** Turf *Poa*	**A8** Prairie	**A9** Turf *Agrostis*	**A10** Turf *Poa*	**B** Mixed
A1											
A2	0.56										
A3	*1.59*	1.59									
A4	0.14	0.05	0.16								
--	0.19	0.11	0.19	0.22							
A6	0.07	0.04	0.08	0.02	0.30						
A7	0.22	0.09	*0.48*	*0.14*	0.22	0.07					
A8	0.31	0.08	0.39	*0.35*	0.36	0.18	*0.49*				
A9	0.27	0.15	0.42	0.45	0.36	0.39	0.74	0.67			
A10	0.05	0.04	0.06	0.05	0.13	0.05	0.10	0.09	0.44		
B	0.51	0.30	0.65	0.67	0.44	0.59	0.73	1.16	0.65	0.32	

Table SI 4.1 (next page)
Colletotrichum cereale strains used in the multi-locus sequence analysis. For lifestyle data: P=pathogen, NP=non-pathogen; and U=pathogenicity unknown, no data. For source data:1=Maria Tomasello-Peterson, University of Mississippi; 2=John Kaminski, University of Connecticut; 3=Frank Wong, University of California Riverside; 4=Tom Hsiang, University of Guelph; 5 =Noel Jackson, University of Rhode Island; 6 =Randy Kane, Chicago Golf District Association; CGD = Dave TeBeest, *Colletotrichum* Germplasm Database, Dept. of Plant Pathology, University of Arkansas; CBS = Centralbureau voor Schimmelcultures, The Netherlands; ICMP = International Collection of Microorganisms from Plants, Auckland, NZ; MAFF = MAFF GenBank, National Institute of Agrobiological Sciences, Tsukuba, Japan.

SI Table 4.1

Group (ID #)	Isolate name	Host plant species	Lifestyle	Country of origin	Other available origination data	Year	Source
A- (21)	NJ-8627	*Poa annua*	P	USA	Manalpan, NJ	2004	--
A1 (1)	1050-AC	*Aegilops cylindrica*	U	USA	Washington Co., AR	1985	CGD 1050
A1 (2)	24049-AS	*Avena sativa*	U	Germany	--	1949	CBS 240.49
A1 (3)	1039-FS	*Festuca* sp.	U	USA	Baldwin Springs, AR	1984	CGD 1039
A1 (4)	1049-HP	*Hordeum pusillum*	U	USA	Washington Co., AR	1985	CGD 1049
A1 (5)	12090-HR	*Hierochloe redolens*	U	New Zealand	Tauranga	1993	ICMP 12090
A2 (10)	KS-TA-10.1A	*Triticum aestivum*	NP	USA	Shawnee Co., KS	2006	--
A2 (11)	KS-FE-7A4	*Festuca elatior*	NP	USA	Tallgrass Prairie Preserve, Cottonwood Falls, KS	2005	--
A2 (12)	KS-20DGU	*Dactylis glomerata*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A2 (12)	KS-20DGY	*Dactylis glomerata*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A2 (12)	KS-20EVD	*Elymus virginicanus*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A2 (12)	KS-20EVM	*Elymus virginicanus*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A2 (12)	KS-20EVX5	*Elymus virginicanus*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A2 (13)	KS-20DGK5	*Dactylis glomerata*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A2 (14)	KS-TA-5W16A	*Triticum aestivum*	NP	USA	Jewell Co., KS	2005	--
A2 (15)	KS-TA-31.1	*Triticum aestivum*	NP	USA	Shawnee Co., KS	2006	--
A2 (16)	510634-AE	*Arrhenatherum elatius*	U	Japan	Chiba Prefecture	1967	MAFF 510634
A2 (6)	305377-AE	*Arrhenatherum elatius*	U	Japan	Chiba Prefecture	1967	MAFF 305377
A2 (7)	KS-TA-32.1.3	*Triticum aestivum*	NP	USA	Shawnee Co., KS	2006	--
A2 (8)	KS-TA-36.2	*Triticum aestivum*	NP	USA	Shawnee Co., KS	2006	--
A2 (9)	NJ-DG1	*Dactylis glomerata*	NP	USA	New Brunswick, NJ	2004	--

SI Table 4.1, cont'd

Group (ID #)	Isolate name	Host plant species	Lifestyle	Country of origin	Other available origination data	Year	Source
A3 (17)	NE-BI-1.3-5	*Bromus inermis*	NP	USA	Willa Cather Prairie, Red Cloud, NE	2005	--
A3 (18)	NE-TA-19.2-A5	*Triticum aestivum*	NP	USA	Pawnee Co. NE	2005	--
A3 (19)	NE-BI-1.2-5	*Bromus inermis*	NP	USA	Willa Cather Prairie, Red Cloud, NE	2005	--
A3 (20)	KS-TA-4-F4	*Triticum aestivum*	NP	USA	Marshall Co., KS	2005	--
A4 (24)	511116-PF	*Polypogon fugax*	U	Japan	Saga Prefecture	1977	MAFF511116
A4 (25)	305429-PF	*Polypogon fugax*	U	Japan	Saga Prefecture	1977	MAFF 305429
A5 (31)	236902-AS	*Agrostis stolonifera*	U	Japan	--	1993	MAFF 238575
A5 (32)	236961-AS	*Agrostis stolonifera*	U	Japan	--	1993	CBS 240.49
A5 (33)	511130-HL	*Holcus lanatus*	U	Japan	Fukushima Prefecture	1977	MAFF 511130
A5 (34)	305384-HL	*Holcus lanatus*	U	Japan	Fukushima Prefecture	1972	MAFF 305394
A5 (35)	305432-HL	*Holcus lanatus*	U	Japan	Saga Prefecture	1977	MAFF 305432
A5 (35)	511140-DG	*Dactylis glomerata*	U	Japan	Tochigi Prefecture	1977	MAFF 511140
A5 (36)	NJ-HL2	*Holcus lanatus*	NP	USA	New Brunswick, NJ	2004	--
A5 (37)	IL-BI-3.5	*Bromus inermis*	NP	USA	Nachusa Prairie, Chicago, IL	2005	--
A5 (38)	305343-HL	*Holcus lanatus*	U	Japan	--	1976	MAFF 305343
A5 (39)	305436-DG	*Dactylis glomerata*	U	Japan	Tochigi Prefecture	1977	MAFF 305436
A5 (40)	305427-AS	*Avena sativa*	U	Japan	Kumamota Prefecture	1977	MAFF 305427
A5 (41)	510530-AS	*Avena sativa*	U	Japan	Saga Prefecture	1966	MAFF 510530
A5 (42)	305371-AS	*Avena sativa*	U	Japan	Saga Prefecture	1966	MAFF 305371
A5 (43)	511114-AS	*Avena sativa*	U	Japan	Kumamoto Prefecture	1977	MAFF 511114
A5 (44)	305075-AS	*Avena sativa*	U	Japan	Saga Prefecture	1966	MAFF 305075
A5 (45)	30569-AA	*Ammophila arenaria*	U	Germany	--	1967	CBS 305.69
A5 (46)	31268-HM	*Holcus mollis*	U	Germany	--	1967	CBS 312.68

SI Table 4.1, cont'd

Group (ID #)	Isolate name	Host plant species	Lifestyle	Country of origin	Other available origination data	Year	Source
A5 (47)	14834-AS	*Avena sativa*	U	Canada	Alberta	1934	CBS 148.34
A5 (48)	305076-AS	*Avena sativa*	U	Japan	Saga Prefecture	1966	MAFF 305076
A6 (28)	IL-CI-7.3D	*Calamagrostis inexpansa*	NP	USA	Indian Boundary Prairie, Markham, IL	2005	--
A6 (30)	KS-10EC1A	*Elymus canadensis*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A6 (30)	KS-10EC2G1	*Elymus canadensis*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A6 (30)	KS-10EC3E2	*Eylmus canadensis*	NP	USA	Konza Prairie, Manhattan, KS	2004	--
A7 (22)	AL-004H11	*Agrostis stolonifera*	P	USA	Birmingham, AL	2006	1
A7 (22)	AL-007132	*Agrostis stolonifera*	P	USA	Birmingham, AL	2006	1
A7 (23)	278CgMA5	*Agrostis stolonifera*	P	USA	Massachusetts	2006	2
A7 (23)	297CgCT7	*Agrostis stolonifera*	P	USA	Connecticut	2006	2
A7 (23)	CA-TCGC575	*Poa annua*	P	USA	Temecula, CA	2002	3
A7 (23)	NC-ABR1	*Poa annua*	P	USA	Monroe, NC	2005	--
A7 (23)	NC-BR11A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR12A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR12B	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR14	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR14A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR18A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR19A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR21B	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR22	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR22B	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR27A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--

SI Table 4.1, cont'd

Group (ID #)	Isolate name	Host plant species	Lifestyle	Country of origin	Other available origination data	Year	Source
A7 (23)	NC-BR28B	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR3B	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR4A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR5	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NC-BR6A	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
A7 (23)	NJ-8626	*Poa annua*	P	USA	Middletown, NJ	2004	--
A7 (23)	TN-GBGC6	*Agrostis stolonifera*	P	USA	Gatlinburg, TN	2005	--
A8 (26)	68188-lg	"lawn grass"	U	Netherlands	--	1988	CBS 681.88
A8 (27)	11021-BW	*Bromus willdenowii*	U	New Zealand	--	1989	ICMP 11021
A9 (52	MS-OW15E302	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (52)	AL-007Z52	*Agrostis stolonifera*	P	USA	Birmingham, AL	2006	1
A9 (52)	AL-007T42	*Agrostis stolonifera*	P	USA	Birmingham, AL	2006	1
A9 (52)	CA-EG20	*Agrostis stolonifera*	P	USA	Corona, CA	2003	3
A9 (52)	NC-ABR26	*Poa annua*	P	USA	Monroe, NC	2005	--
A9 (52)	NC-ABR44	*Poa annua*	P	USA	Monroe, NC	2005	--
A9 (52)	NC-ABR52	*Poa annua*	P	USA	Monroe, NC	2005	--
A9 (52)	NC-ABR14	*Poa annua*	P	USA	Monroe, NC	2005	--
A9 (52)	NC-ABR7	*Poa annua*	P	USA	Monroe, NC	2005	--
A9 (52)	MS-OW15	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (52)	MS-OW15D21	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (52)	MS-OW15H1A2	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (52)	MS-OW15R13	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (52)	MS-OW15R41	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (52)	TN-GBGC3	*Agrostis stolonifera*	P	USA	Gatlinburg, TN	2005	--

SI Table 4.1, cont'd

Group (ID #)	Isolate name	Host plant species	Lifestyle	Country of origin	Other available origination data	Year	Source
A9 (52)	TN-GBGC4	*Agrostis stolonifera*	P	USA	Gatlinburg, TN	2005	--
A9 (52)	TN-GBGC554	*Agrostis stolonifera*	P	USA	Gatlinburg, TN	2005	--
A9 (52)	VA-PA5	*Agrostis stolonifera*	P	USA	Virginia Beach, VA	2005	--
A9 (53)	ONT-00170	*Agrostis stolonifera*	P	Canada	Downsville, ON	2000	4
A9 (54)	NJ-HF2B	*Poa annua*	P	USA	New Brunswick, NJ	2003	--
A9 (54)	TN-GBGC5	*Agrostis stolonifera*	P	USA	Gatlinburg, TN	2006	--
A9 (54)	VA-PA1	*Agrostis stolonifera*	P	USA	Virginia Beach, VA	2005	--
A9 (55)	ONT-99359	*Agrostis stolonifera*	P	Canada	Victoria East, Guelph, ON	1999	4
A9 (56)	NJ-7130	*Poa annua*	P	USA	Atlantic Co., NJ	2004	--
A9 (57)	NY-8893	*Agrostis stolonifera*	P	USA	Scarsdale, NY	2006	--
A9 (57)	MS-OW15F52	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (57)	MS-OW15H32	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
A9 (58)	VA-PA3	*Agrostis stolonifera*	P	USA	Virginia Beach, VA	2005	--
A9 (59)	VA-PA2	*Agrostis stolonifera*	P	USA	Virginia Beach, VA	2005	--
A9 (60)	CT-14	*Poa annua*	P	USA	Aspetuck Valley, CT	1998	5
A9 (81)	MA-24	*Agrostis stolonifera*	P	USA	Pine Acres, MA	1998	5
A10 (61)	KS-20BIG	*Bromus inermis*	NP	USA	Konza Prairie, Manhattan,KS	2004	--
A10 (62)	NJ-6795	*Poa annua*	P	USA	Somerset Co., NJ	2003	--
A10 (63)	NY-8422	*Poa annua*	P	USA	Rye, NY	2004	--
A10 (64)	NJ-6340	*Poa annua*	P	USA	Ocean Co., NJ	2003	--
A10 (65)	ME-3	*Poa annua*	P	USA	Kebo Valley, ME	1998	5
A10 (66)	MA-11	*Poa annua*	P	USA	Worchester, MA	1998	5
A10 (67)	IL-PV2	*Poa annua*	P	USA	Chicago, IL	2005	6
A10 (68)	CA-SC44	*Poa annua*	P	USA	San Jose, CA	2003	3

SI Table 4.1, cont'd

Group (ID #)	Isolate name	Host plant species	Lifestyle	Country of origin	Other available origination data	Year	Source
A10 (69)	CA-SH29	*Poa annua*	P	USA	San Bernadino, CA	2003	3
A10 (70)	CA-FUGC1143	*Poa annua*	P	USA	Fullerton, CA	2004	3
A10 (71)	CA-ANCG1715	*Poa annua*	P	USA	Pasadena, CA	2004	3
A10 (72)	CA-FUGC1	*Poa annua*	P	USA	Fullerton, CA	2003	3
A10 (73)	CA-EG15	*Poa annua*	P	USA	Corona, CA	2003	4
A10 (74)	CA-AHCC1049	*Poa annua*	P	USA	Arcadia, CA	2005	3
B (75)	PA-50181	*Poa annua*	P	USA	Reedsville, PA	2000	--
B (77)	RI-22	*Agrostis stolonifera*	P	USA	Metacomet, RI	1998	5
B (78)	MS-OW15P51	*Agrostis stolonifera*	P	USA	West Point, MS	2006	1
B (79)	NC-ABR48	*Poa annua*	P	USA	Monroe, NC	2005	--
B (82)	ONT-00128	*Agrostis stolonifera*	P	Canada	Osprey Valley, ON	2000	4
B (83)	NJ-DG-2A2-2	*Dactylis glomerata*	NP	USA	Walkill WMA, Sussex County, NJ	2005	--
B (84)	ALB-99325	*Poa pratensis*	P	Canada	Lacombe, AL	1999	4
B (85)	MA-6722	*Poa annua*	P	USA	Massachusetts	2003	--
B (86)	NC-BR27B	*Poa annua*	P	USA	Blowing Rock, NC	2005	--
B (87)	PA-50005	*Poa annua*	P	USA	Bernville, PA	1998	--
B (88)	NJ-6607	*Poa annua*	P	USA	Morris Co., NJ	2003	--
B (89)	NJ-6491	*Poa annua*	P	USA	Morris Co., NJ	2003	--
B (90)	NJ-DG-2A2-5	*Dactylis glomerata*	NP	USA	Walkill WMA, Sussex County, NJ	2005	--
B (91)	NJ-CA1C1	*Calamagrostis acutifolia*	P	USA	Barrington, NJ	2005	--
B (92)	NJ-CA1N2	*Calamagrostis acutifolia*	P	USA	Barrington, NJ	2005	--
B (93)	30469-AA	*Ammophila arenaria*	U	Germany	--	1967	CBS 304.69
B (94)	305369-AT	*Agrostis tenuis*	U	Germany	--	1967	CBS 303.69
B (96)	NJ-DG-4.4	*Dactylis glomerata*	NP	USA	New Brunswick, NJ	2005	--

SI Table 4.2 (next page)

Isolates of grass-associated *Colletotrichum* species and outgroup taxa used in this study. For source data:1=Lisa Vaillancourt, University of Kentucky; 2=Gary Bergstrom, Cornell University; 3=Peter Oudemans, Rutgers University; IMI= CABI Genetic Resource Collection, Surrey, UK; CGD = Dave TeBeest, *Colletotrichum* Germplasm Database, Dept. of Plant Pathology, University of Arkansas; CBS = Centralbureau voor Schimmelcultures, The Netherlands; ICMP = International Collection of Microorganisms from Plants, Auckland, NZ; MAFF = MAFF GenBank, National Institute of Agrobiological Sciences, Tsukuba, Japan.

SI Table 4.2.

Isolate name	Host plant species	Species	Country of origin	Other available origination data	Year	Source
1040-DS	*Digitaria* sp.	*C*. ex. digitaria	USA	Monticello, AR	1992	CGD 1040
1047-PD	*Paspalum dilatatum*	*C*. ex. paspalum	USA	Baldwin Springs, AR	1985	CGD 1047
1391-DS	*Digitaria* sp.	*C*. ex. digitaria	USA	--	--	CGD 1391
16970-SO	*Saccharum officinarum*	*C*. ex. saccharum	Brazil	--	1970	CBS 169.70
176617-BB	*Bothriochloa bladhii*	*C. caudatum*	Australia	--	1973	IMI 176617
176619-IC	*Imperata cylindrica*	*C. caudatum*	Australia	Caboolture	1973	IMI 176619
238575-ZT	*Zoysia tennufolia*	*C*. ex. digitaria	Japan	--	--	MAFF 238575
24362-SO	*Saccharum officinarum*	*C. falcatum*	Brazil	--	1962	CBS 243.62
27554-ZM	*Zea mays*	*C. graminicola*	Netherlands	--	1954	CBS 275.54
279189-AA	*Axoponus affinis*	*C*. ex. axonponus	Australia	Queensland	1983	IMI279189
305077-SO	*Saccharum officinarum*	*C*. ex. saccharum	Japan	Chiba Prefecture	1966	MAFF 305077
305360-SB	*Sorghum bicolor*	*C. sublineolum*	Japan	--	1957	MAFF 305360
305361-SB	*Sorghum bicolor*	*C. sublineolum*	Japan	--	1957	MAFF 306261
305391-PD	*Paspalum dilatatum*	*C*. ex. paspalum	Japan	Chiba Prefecture	1974	MAFF 305391
305403-PN	*Paspalum notatum*	*C*. ex.. paspalum	Japan	Yamaguchi Prefecture	1977	MAFF 305403
305404-DC	*Digitaria ciliaris*	*C*. ex. digitaria	Japan	Tochigi Prefecture	1975	MAFF 305404
305428-PN	*Paspalum notatum*	*C*. ex. paspalum	Japan	Kumamato Prefecture	1977	MAFF 305428
305439-EE	*Echinochloa esculenta*	*C*. ex. echinochloa	Japan	Miyazaki Prefecture	1977	MAFF 305439
305460-EE	*Echinochloa esculenta*	*C*. ex. echinochloa	Japan	Tochigi Prefecture	1980	MAFF 305460
305700-IC	*Imperata cylindrica*	*C. caudatum*	Japan	--	--	MAFF 305700
306170-SO	*Saccharum officinarum*	*C. falcatum*	Japan	--	--	MAFF 306170
306299-SO	*Saccharum officinarum*	*C. falcatum*	Japan	--	--	MAFF 306299
311343-ZM	*Zea mays*	*C. graminicola*	Japan	--	1985	MAFF 311343

SI Table 4.2, cont'd.

Isolate name	Host plant species	Species	Country of origin	Other available origination data	Year	Source
347765-SO	*Saccharum officinarum*	*C. falcatum*	Nigeria	--	--	IMI 347765
510020-SB	*Sorghum bicolor*	*C. sublineolum*	Japan	--	1957	MAFF 510020
510021-SB	*Sorghum bicolor*	*C. sublineolum*	Japan	--	1957	MAFF 510021
510857-MS	*Miscanthus sinensis*	*C.* ex. miscanthus	Japan	Tochigi Prefecture	1972	MAFF 510857
510916-PD	*Paspalum dilatatum*	*C.* ex. paspalum	Japan	Chiba Prefecture	1975	MAFF 510916
511000-PN	*Paspalum notatum*	*C.* ex. paspalum	Japan	--	1975	MAFF 511000
511014-DC	*Digitaria ciliaris*	*C.* ex. digitaria	Japan	Tochigi Prefecture	1975	MAFF 511014
511115-PD	*Paspalum dilatatum*	*C.* ex. paspalum	Japan	--	--	MAFF 511115
511152-EE	*Echinochloa esculenta*	*C.* ex. echinochloa	Japan	Kochi Prefecture	1977	MAFF 511152
511155-EI	*Eleusine indica*	*C.* ex. eleusine	Japan	Kumamoto Prefecture	1977	MAFF 511155
511328-EE	*Echinochloa esculenta*	*C.* ex. echinochloa	Japan	Tochigi Prefecture	1980	MAFF 511328
511344-EE	*Echinochloa esculenta*	*C.* ex. echinochloa	Japan	Tochigi Prefecture	1985	MAFF 511344
5171-PS	*Paspalum* sp.	*C.* ex. paspalum	New Zealand	Auckland	1965	ICMP 5171
78362-SO	*Saccarum officinarum*	*C. falcatum*	Nigeria	--	1960	IMI 78362
M5001	*Zea mays*	*C. graminicola*	Brazil	--	1990	1
IN-10472	*Zea mays*	*C. graminicola*	USA	Benton Co., IN	1972	1
IN-10570	*Zea mays*	*C. graminicola*	USA	Indiana	1970	1
IN-10670	*Zea mays*	*C. graminicola*	USA	Indiana	1970	1
IN-10970	*Zea mays*	*C. graminicola*	USA	Indiana	1970	1
IN-12270	*Zea mays*	*C. graminicola*	USA	Indiana	1970	1
IN-12475	*Zea mays*	*C. graminicola*	USA	Indiana	1975	1
IN-300170	*Zea mays*	*C. graminicola*	USA	Indiana	1970	1
IN-900190	*Zea mays*	*C. graminicola*	USA	Indiana	1990	1

SI Table 4.2, cont'd.

Isolate name	Host plant species	Species	Country of origin	Other available origination data	Year	Source
IN-D77	*Zea mays*	*C. graminicola*	USA	Indiana	1977	1
IN-DUB90	*Zea mays*	*C. graminicola*	USA	Dubois Co., IN	1990	1
KY-197	*Zea mays*	*C. graminicola*	USA	Mc Clean Co, KY	1997	1
KY-297	*Zea mays*	*C. graminicola*	USA	Union Co. KY	1997	1
KY-397	*Zea mays*	*C. graminicola*	USA	Fayette Co.KY	1997	1
KY-398	*Zea mays*	*C. graminicola*	USA	Kentucky	1998	1
M1001	*Zea mays*	*C. graminicola*	USA	Missouri	1978	1
MO-178	*Zea mays*	*C. graminicola*	USA	Missouri	1978	1
MO-478	*Zea mays*	*C. graminicola*	USA	Missouri	1978	1
MO-878	*Zea mays*	*C. graminicola*	USA	Missouri	1978	1
MO-978	*Zea mays*	*C. graminicola*	USA	Missouri	1978	1
NC(-200170	*Zea mays*	*C. graminicola*	USA	North Carolina	1970	1
NY-15182	*Zea mays*	*C. graminicola*	USA	Tioga Co., NY	1982	2
NY-400180	*Zea mays*	*C. graminicola*	USA	New York	1980	1
NY-AU80	*Zea mays*	*C. graminicola*	USA	Cayuga Co. NY	1980	1
S17001	*Sorghum bicolor*	*C. sublineolum*	Texas	--	--	1
S19001	*Sorghum bicolor*	*C. sublineolum*	South Africa	--	--	1
S3001	*Sorghum bicolor*	*C. sublineolum*	Burkina Fasso	--	--	1
TX-BI2K	*Sorghum halapense*	*C. sublineolum*	USA	Peach Point WMA, Brazoria Co., TX	2005	--
PA-3AP	*Malus domestica*	*C. acutatum*	USA	Pennsylvania	2004	3
VA1-AP	*Malus domestica*	*C. acutatum*	USA	Virginia	2003	3

Figure 4.2 -- Next page.

Legend. The maximum likelihood phylogenetic tree topology constructed from the combined four gene 3031-bp dataset, illustrating the evolutionary relationships between the graminicolous *Colletotrichum*. The tree was rooted using outgroup taxa from *C. acutatum* (not shown; SI Table 2). Only a representative sample of *C. cereale* isolates are included in the tree for clarity and because the presence of recombination in the dataset showed that these isolates were not accurately depicted within a bifurcating tree topology. The phylogenetic prediction of the complete *C. cereale* topology can be viewed with SI Figure 5. Bayesian posterior probabilities supporting the topology are shown below the branches; support values below 50 are not labeled.

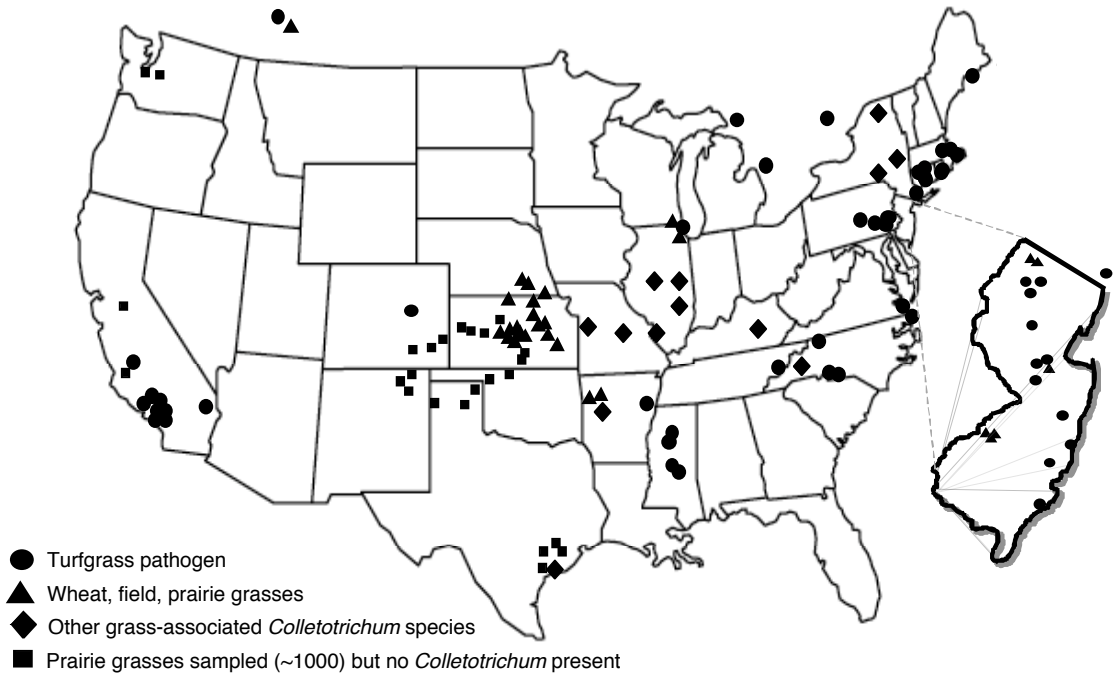


Figure 4.1
North American collection sites sampled in this study (SI Tables 1 and 2). *C. cereale* was also collected from Japan, Germany, New Zealand and The Netherlands (SI Table 1).

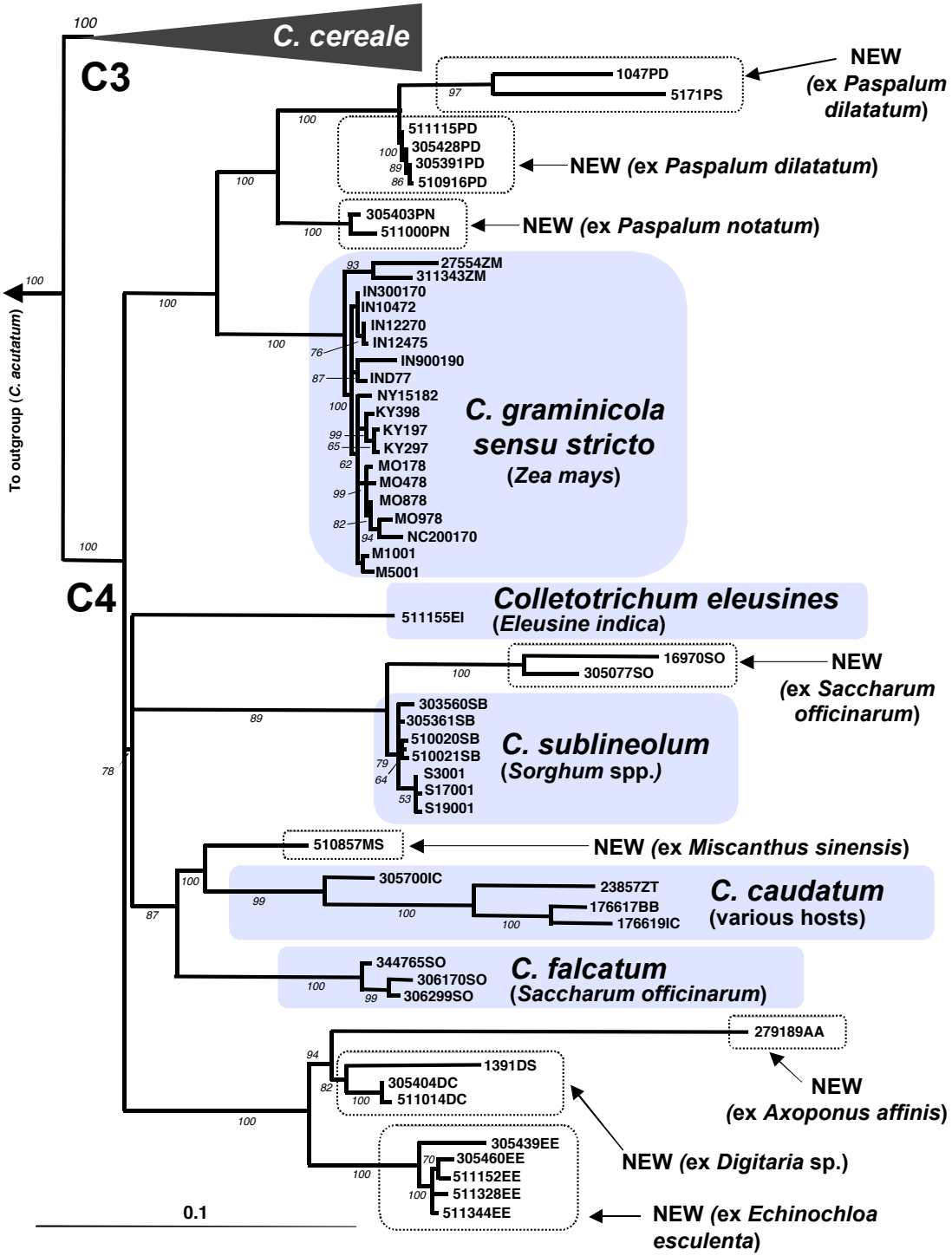


Figure 4.2

The maximum likelihood phylogenetic tree topology constructed from the combined four gene 3031-bp dataset, illustrating the evolutionary relationships between the graminicolous *Colletotrichum*. The tree was rooted using outgroup taxa from *C. acutatum* (not shown; SI Table 4.2). Bayesian posterior probabilities supporting the topology are shown below the branches; support values below 50 are not labeled.

Figure 4.3 - shown on next page.
C. cereale median joining network. Populations containing international and/or historical isolates (pre-1998) are labeled (A1, A2, A3, A4, A5, A6, A8 and B); populations comprised exclusively of modern North American (USA and Canada, 1998 to present) turfgrass-derived populations are unlabeled (A7, A9 and A10). Pathogenic haplotypes are represented by circles, non-pathogen haplotypes are represented by hexagons, and haplotypes of unknown pathogenicity are represented by squares. Color coding of haplotypes corresponds to ecosystem and/or host, as illustrated in the key.

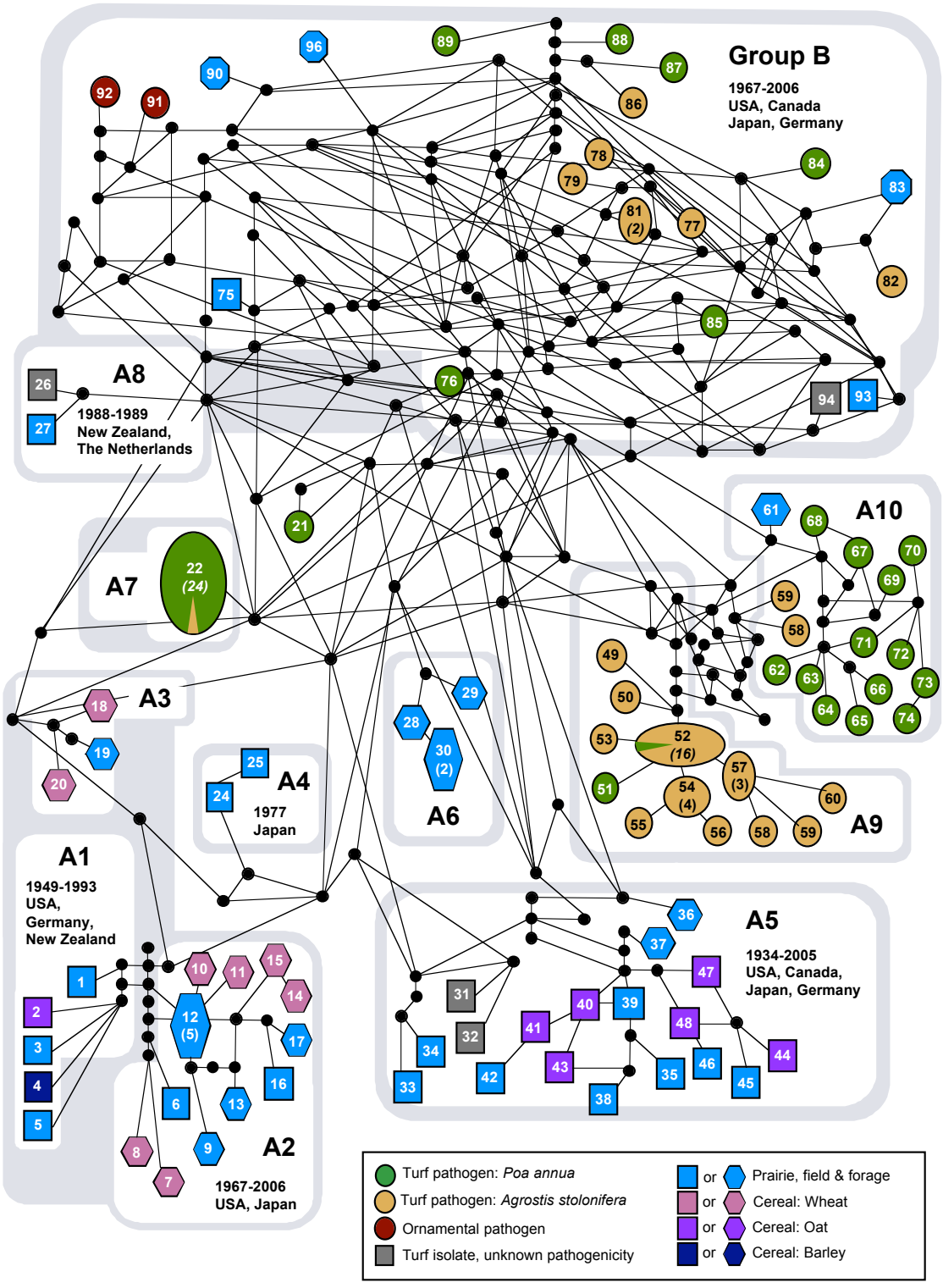


Figure 4.3. Legend on previous page.

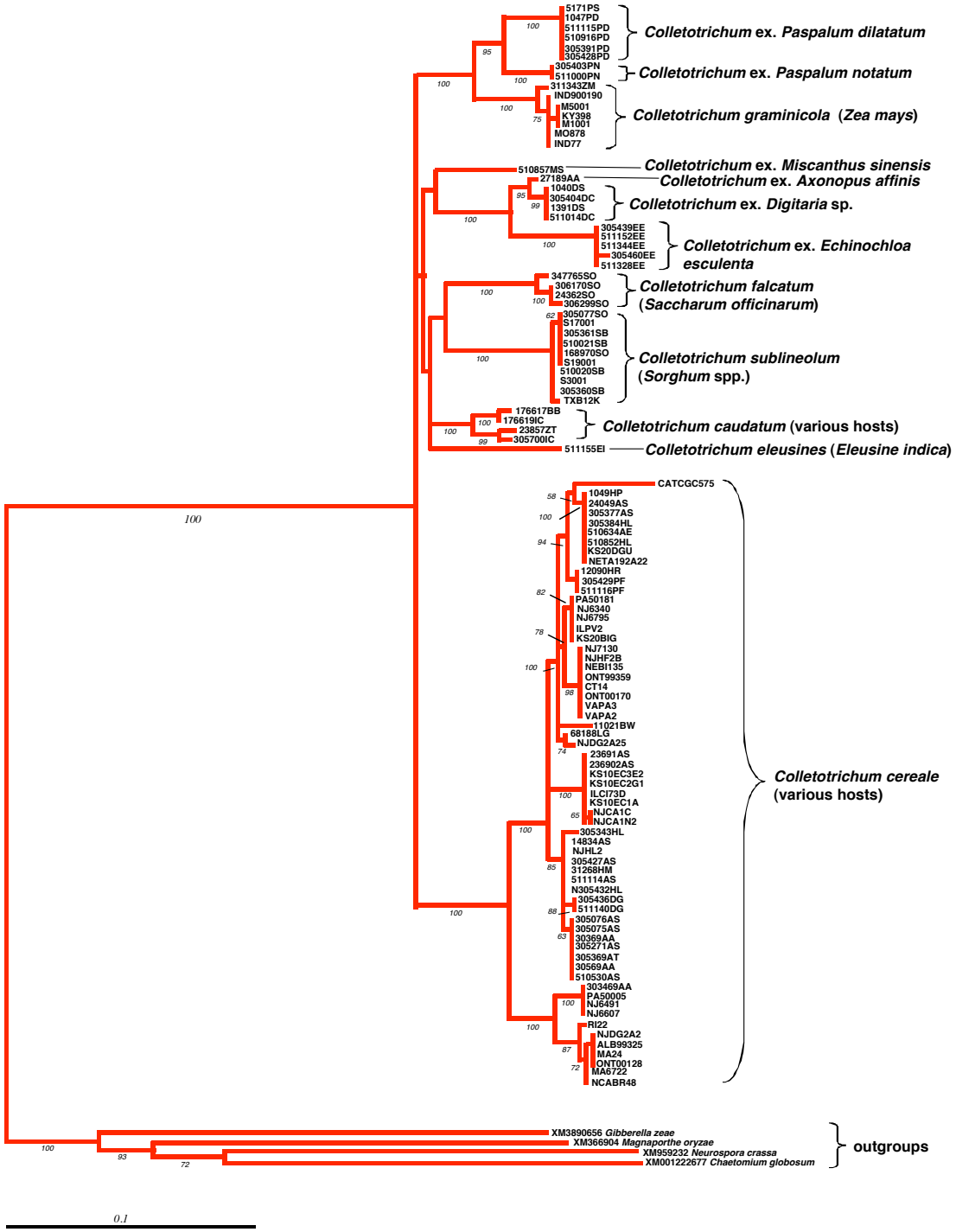


Figure 4.4

Maximum likelihood derived gene tree depicting the relationships between the grass inhabiting species of *Colletotrichum*, constructed using the 850-bp W1 dataset, which corresponds to 5' end of the DNA lyase gene adjacent to the *Mat1* locus. Bayesian posterior probabilities supporting the topology are shown below the branches; support values below 50 are not labeled.

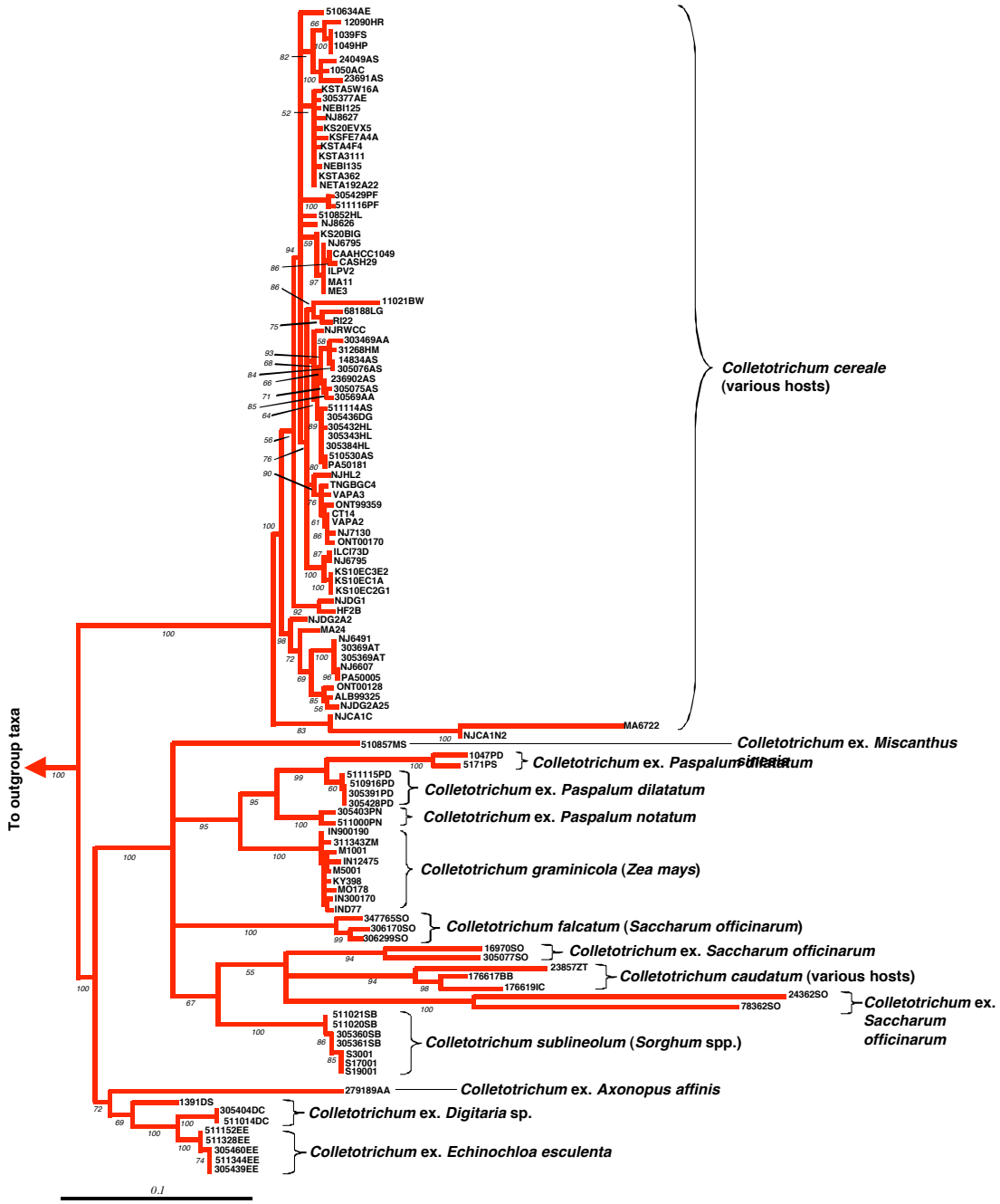


Figure 4.5

Maximum likelihood derived gene tree depicting the relationships between the grassinhabiting species of *Colletotrichum*, constructed using the 1,400-bp M72 dataset, which corresponds tothe 3' end of the DNA lyase gene (*Apn1*) adjacent to the *Mat1* locus, the intergenic region between *Apn1* and *Mat1*, and the 3' end of *Mat1*. Bayesian posterior probabilities supporting the topology are shown below the branches; support values below 50 are not labeled.

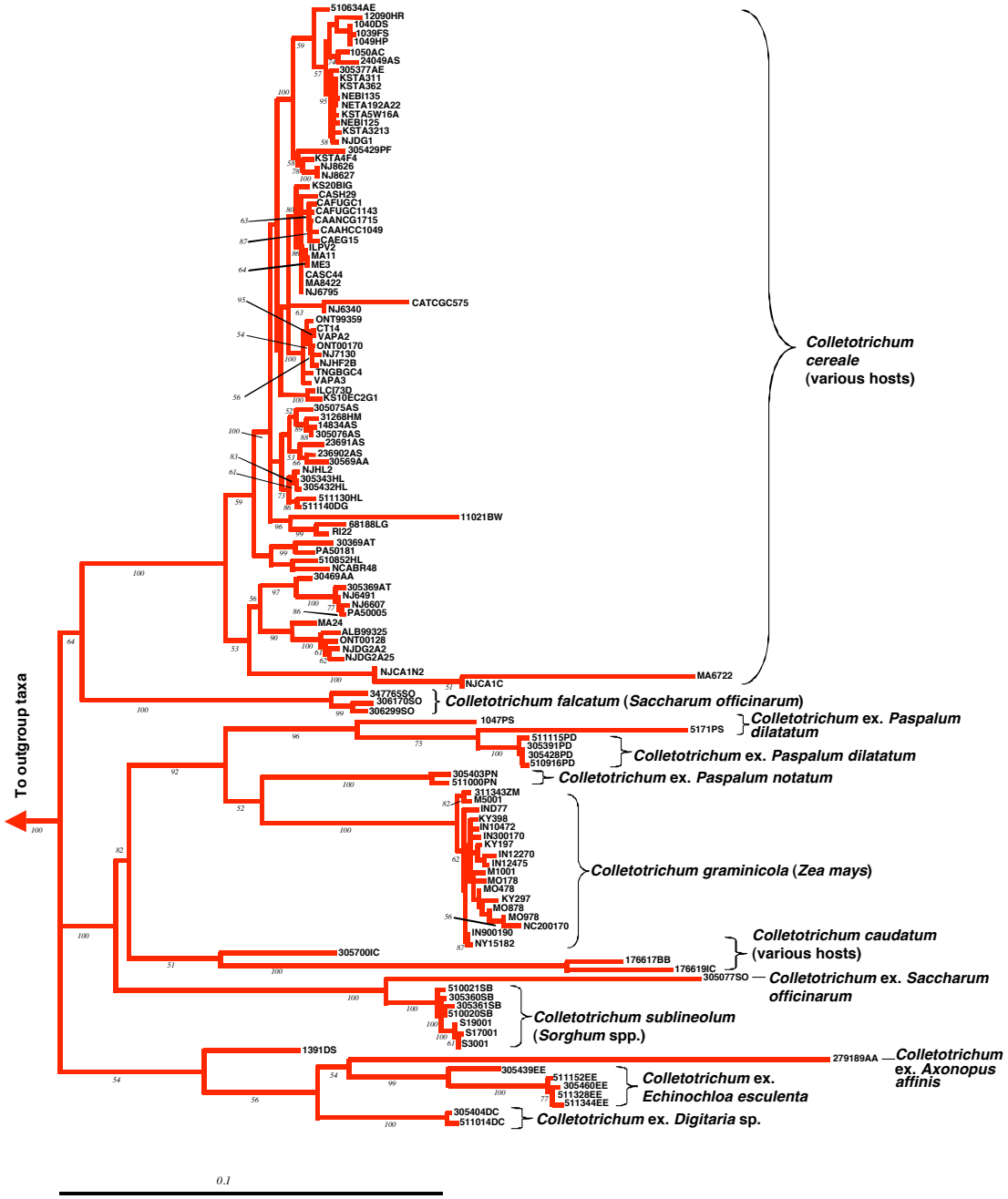


Figure 4.6

Maximum likelihood phylogenetic tree constructed using the550-bp *Sod2* gene sequence. Bayesian posterior probabilities supporting the tpology are shown below the branches; support values below 50 are not labeled.

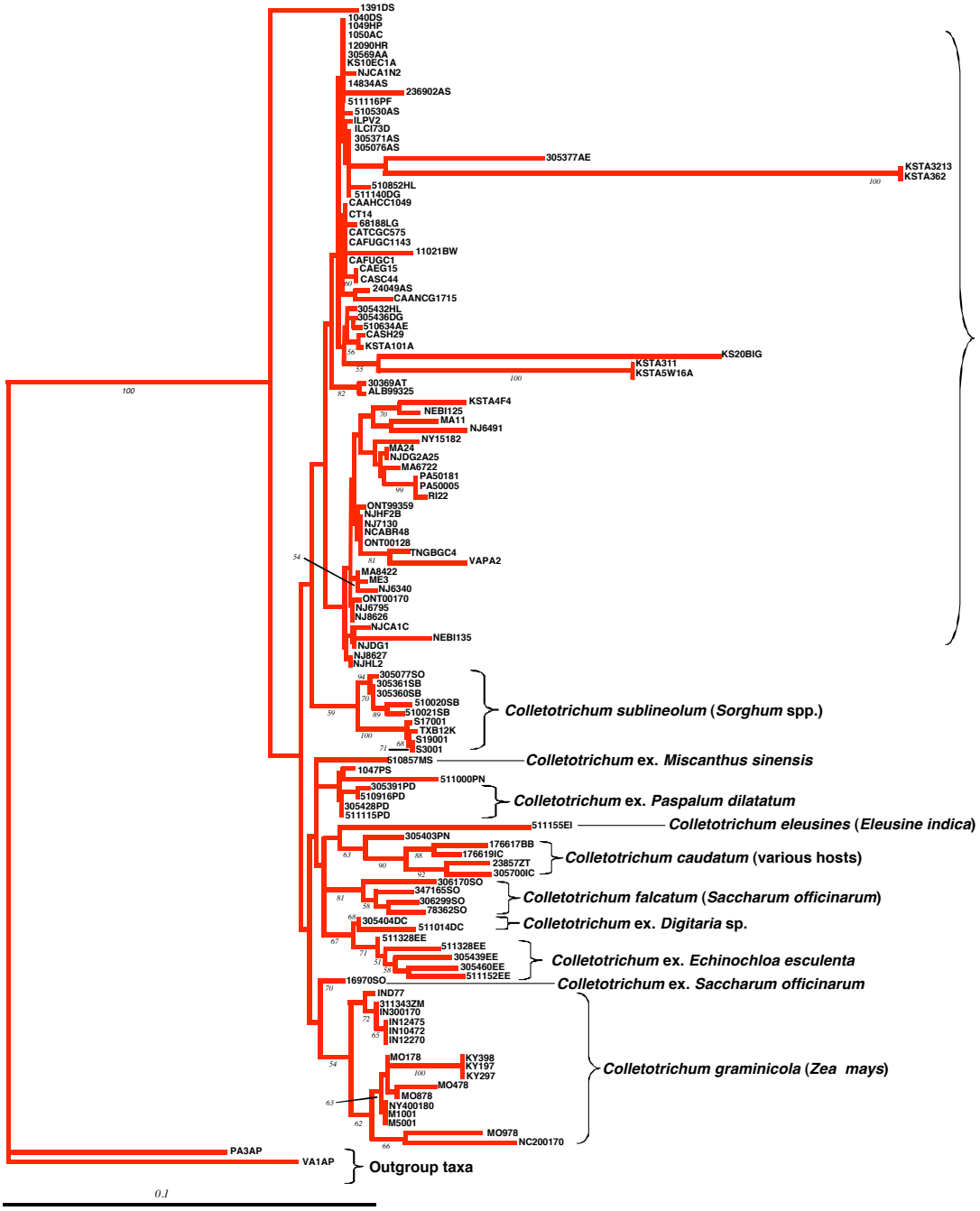


Figure 4.7

Maximum likelihood phylogenetic tree constructed using the 450-bp ITS region. Bayesian posterior probabilities supporting the topology are shown below the branches; support values below 50 are not labeled.

Figure 4.8

Maximum likelihood phylogenetic tree of *Colletotrichum cereale* isolates using the combined dataset. Bayesian posterior probabilities supporting the tpology are shown below the branches; support values below 50 are not labeled.

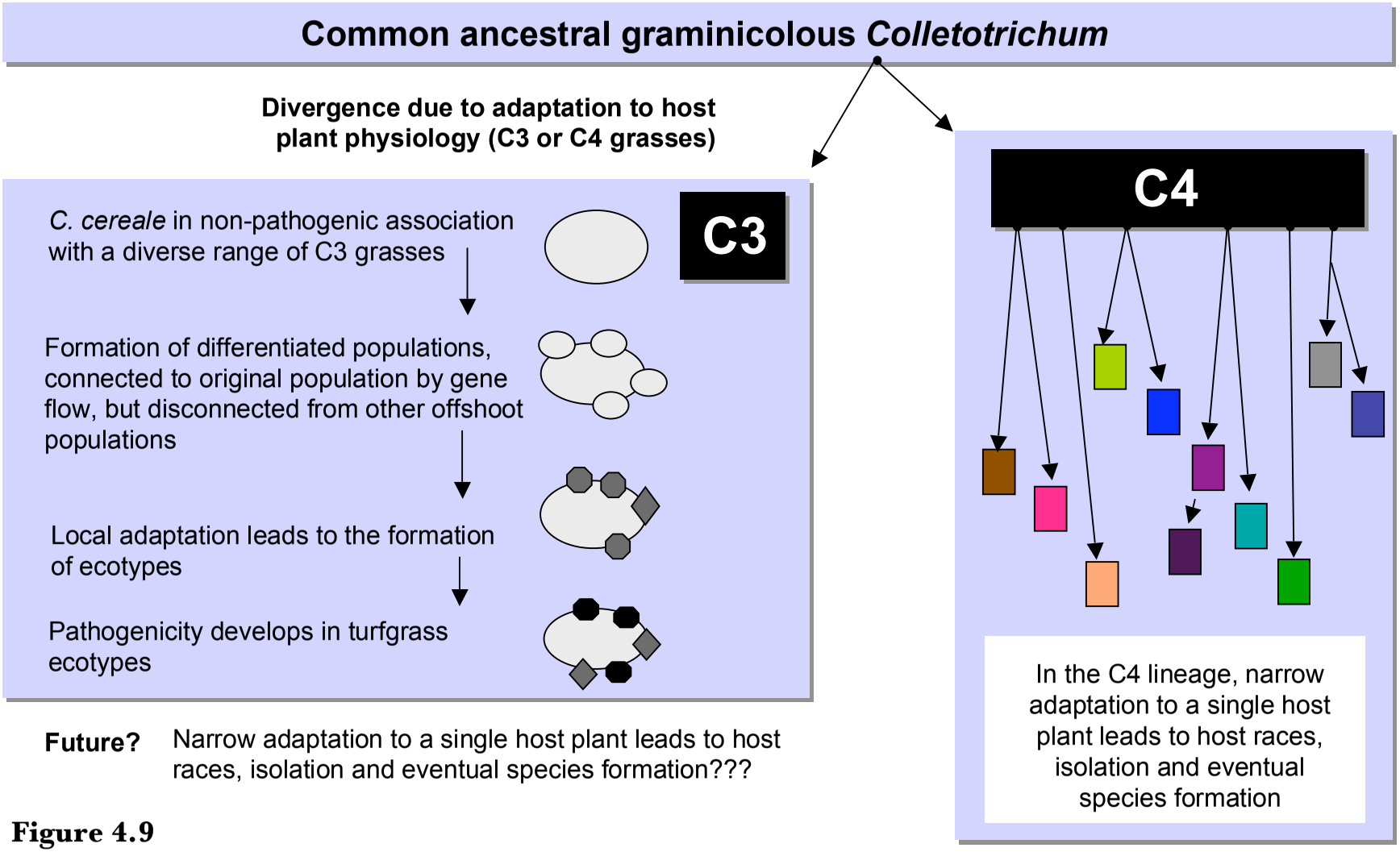


Figure 4.9
Summary figure illustrating a hypothesis of the evolutionary history of the falcate-spored gramnicolous *Colletotrichum* suggested by this research.

Chapter 5: The development of microsatellite markers as tools to expand population analyses of *Colletotrichum cereale* in turfgrass, cereal crops and natural ecosystems

ABSTRACT

In this chapter, I describe the isolation and preliminary characterization of 22 microsatellite loci from the fungus *Colletotrichum cereale* suitable for use in fine-scale population genetic analyses of this economically important pathogen of cool-season (C3) grasses. These sequences originated from a plasmid DNA library enriched for dinucleotide repeats; they are the first microsatellite markers developed for any member of this genus. Polymerase chain reaction (PCR) for 22 of the 35 total microsatellite sequences produced polymorphic PCR profiles from 10-100-bp range from 12 *C. cereale* samples drawn from 11 characterized populations. Although 25 of the 35 loci are widely distributed across all 11 *C. cereale* populations, only 10 of the loci are present in the whole genome shotgun sequence of the sister taxon *C. graminicola*, and just one of the *C. graminicola* loci possess the repeat motifs. Given their polymorphic profiles across the species range, these microsatellite markers will be useful in future studies of *C. cereale* populations, and contribute to the growing number of genomic resources available for this increasingly important evolutionary and ecological research system.

5.1 Introduction

In recent years the fungus *Colletotrichum cereale* has emerged as a commercially important pathogen of cool-season turfgrasses, causing destructive outbreaks of anthracnose disease in stands of *Poa annua* and *Agrostis stolonifera* maintained as golf course greens across North America (Crouch *et al.*, 2005; Crouch *et al.*, 2006; Crouch *et al*, 2008a; Smiley *et al*, 2005; Chapter 1; Chapter 4) and the United Kingdom (Mann and Newell, 2005). Outside of the cultivated turfgrass environment, the fungus is noteworthy in that it is a very common, typically non-pathogenic inhabitant of a wide range of cool-season grasses across a diversity of ecosystems including cereal crops, prairie and other native grasslands, weedy grasses and field and forage grasses (Crouch *et al.,* 2006; Crouch *et al.,* 2008a; chapter 1; chapter 4). Phylogenetic and population analyses of the fungus isolated from diseased and asymptomatic grasses using multi-locus nucleotide sequence analysis identified a predominance of *C. cereale* populations that were structured almost entirely according to lifestyle; i.e., pathogens of turfgrass or non-pathogenic inhabitants of non-turfgrass environments (Crouch *et al.,* 2008a; chapter 4). This pattern of population subdivision was observed irrespective of geographic origin, with 91% of the *C. cereale* populations comprised entirely of either turfgrass pathogens or non-pathogenic isolates (Crouch *et al.,* 2008a; chapter 4). The three observed primary turfgrass pathogen populations were highly specialized, with populations divided based upon host plant origin: two populations were primarily derived from *P. annua* and one primarily from *A. stolonifera* (Crouch *et al.,* 2008a; chapter 4). In order to evaluate the differences that make the turfgrass pathogen isolates and agronomic/prairie isolates of *C. cereale* unique from one another, it is necessary to evaluate *C. cereale* populations on a finer scale, both with

respect to sampling and the genetic markers employed. Given the variability in effectiveness of anthracnose disease management strategies at different golf course sites, higher resolution genotyping of *C. cereale* isolates was sought to provide answers to epidemiological questions important for effective disease control on golf course turf: inoculum sources and routes of transmission, the genotypic makeup of isolates and their association e with pathogenicity to different grass species, and the recent origin and initiation of the turfgrass anthracnose disease epidemics. Microsatellite markers are ideal for these tasks. Highly polymorphic and abundant in the genomes of most eukaryotes, microsatellite sequences consist of short, tandemly repeated units comprised of two, three, four or more nucleotides, with polymorphisms arising through replication slippage or mutations extending or interrupting a series of repeats. In fungal populations, microsatellite motifs have been shown to be highly variable, with little evidence of intraspecific saturation, and mutation rates exceeding those experienced by flanking DNA by ~2500x (Dettman and Taylor, 2004). In empirical studies of fungal plant pathogens, microsatellite markers have provided the tools to characterize host related genetic differentiation, field-scale population structure, founder effects, sexual reproduction, and dispersal of plant pathogens (e.g., van Putten *et al*, 2005; Raboin *et al.*, 2007; Rivas *et al.*, 2004; Sexton *et al.*, 2006; Stuckenbrock *et al.*, 2006). Here I report the isolation and characterization of 22 polymorphic microsatellite loci from the genome of *C. cereale* that are suitable for use as rapid, cost-effective, reliable, and comprehensive genotypic markers and that will facilitate future studies of *C. cereale* populations.

5.2 Materials and methods

A genomic DNA library enriched for dinucleotide repeat sequences (the most abundant class of microsatellites found in eukaryotic genomes) was constructed following a modification of the protocol described by Carlton *et al.,* (2002). Briefly, genomic DNA was extracted from *C. cereale* isolate NJ-6340 using a standard phenol-chloroform purification as described previously (Crouch *et al.*, 2005). DNA was digested to completion using *Sau3A*I and purified on a 1.2% agarose gel. DNA fragments of 400-900-bp were excised from the gel and purified with the GeneClean III kit following the manufacturers protocol (Qbiogene, Irvine, CA) was and were ligated to linker DNA constructed from the oligonucletides A (5'GATCGTCGACGGTACCGAATTCT) and B (5'GTCAAGAATTCGGTACCGTCGAC) (synthesized by Sigma Genosys, The Woodlands, TX). After column purification with the Pure-Link PCR purification kit (Invitrogen Corp., Carlsbad, CA), 10 cycles of PCR using oligonucleotide A were performed to amplify the linker-ligated fragments and to select DNA with linkers on both sides of the insert and increase the number of these fragments. These amplification products were column purified and enriched for microsatellite sequences by hybridizing of a biotinylated oligonucleotide probe (5'Biotin-ATAGAATAT$[CA]_{16}$) (synthesized by Sigma Genosys, The Woodlands, TX) with the DNA, followed by purification using Dynal Streptavidin Magnetic Dynabeads (Invitrogen, Inc., Carlsbad, CA). After enrichment, DNA was PCR amplified using primer A, column purified, and cloned into the pGEM-T Easy vector following the manufacturers protocol (Promega, Madison, WI). Recombinant plasmids were introduced into *Escherichia coli* DH5α competent cells (Invitrogen Corp., Carlsbad, CA) and resulting colonies were screened for inserts.

Plasmids for sequencing were purified using the QIAprep Spin Miniprep Kit (Qiagen, Inc., Valencia, CA) and sequenced in both directions using universal T7 and SP6 primers and ABI Big Dye 3 chemistry (Applied Biosystems Inc., Foster City, CA) at one quarter the manufacturer's recommendation. Sequences were assembled using the SeqMan program (DNASTAR, Madison, WI).

PCR primers flanking the microsatellite motifs were designed using the Primer3Plus interface (www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi; Kõressaar and Romm, 2007). PCR primer pairs were designed and synthesized for microsatellite-containing sequences to amplify DNA sized between100-300-bp (Table 5.1). In the present study size determination was performed entirely using agarose gel electrophoretic analysis, but future implementation of these primer pairs will rely upon automated laser detection of fluorescent-labelled PCR product. To facilitate automated fluorescence-based detection, PCR primer design was guided by two basic criteria.: first, primer pairs were optimized with an optimal annealing temperature of 55°C, and second, the universal M13 sequence (5'- GGAAACAGCTATGACCAT-3'; 53°C optimal annealing temperature) was added to the 5' end of the forward primer. This standardization of primers will permit the implementation of a cost effective, two-round amplification procedure (Schuelke, 2000) in later population genotyping, allowing the expensive fluorescent dyes to be attached to a universal M13 sequence rather than individually modifying each microsatellite marker.

Microsatellite sequences were PCR amplified in 15 μl reactions containing 22.5 ng of genomic DNA, 0.6 U Chroma-Taq DNA polymerase (Denville Scientific, Inc., Metuchen, NJ) in 10x PCR buffer, 1.5 mM of $MgCl_2$, 0.2 mM of each dNTP, and 7.5 ng of each primer. Samples were amplified as follows: 95°C for 5 min, then 40 cycles of 30 s

at 95°C, 45 s at 55°C, 45 at 72°C and a final 72°C extension for 10 min. Negative controls were included in all amplifications to check for possible contamination, and a reaction containing genomic DNA of library isolate NJ-6340 was included as a positive control. Negative reactions were repeated twice with appropriate controls to confirm absence of amplification product. Amplified fragments were visualized on 1.5% agarose gels run in 0.5% TBE buffer and stained with ethidium bromide.

The microsatellite sequences generated in this study were submitted to the NCBI GenBank and assigned accession numbers EU554365–EU554396. An embargo was placed on release of the sequences through March 10, 2009 or upon publication of this research, whichever comes first. Cloned inserts containing non-microsatellite sequences were submitted to GenBank in association with a separate study (Crouch *et al*, 2008b; chapter 2).

5.3 Results and discussion

A total of 123 insert-containing colonies were sequenced and evaluated for the presence of microsatellite motifs. Insert sizes ranged from 122-bp to 759-bp. Numerous colonies contained dinucleotide, or less frequently, trinucleotide repeats; however, flanking sequence suitable for PCR priming was present only in 35 of the sequences (Fig. 5.1). No high quality matches with curated sequences were identified through BLAST X searches of the NCBI GenBank database when the 35 microsatellite-containing sequences were used as queries. Based on these results, it appears that the microsatellite-containing sequences are derived from non-coding intergenic or intronic regions rather than protein coding genes.

PCR reactions using the 35 primer pairs designed using the flanking regions of the microsatellite-containing loci were used to screen a representative sample of *C. cereale* isolates (Table 5.2) drawn from the eleven major populations of this fungus previously characterized using multi-locus nucleotide sequence analysis (Crouch *et al*, 2008a; chapter 4). Table 5.3 shows the results of these reactions, as visualized on a 1.5% TBE agarose gel. Of the 35 primer pairs, 26 (74%) resulted in amplification products from at least one *C. cereale* isolate. No product was observed from the remaining nine primer pairs even from the genomic DNA of the template isolate NJ-6340 (not shown). Because one of the objectives in the development of these microsatellites is reproducible amplification using 55^{o} C annealing temperatures, no attempt was made to optimize annealing temperatures using gradient PCR cycling, and no further work was performed using these nine sequences.

A distinct pattern of presence or absence at the 26 amplifiable microsatellite-containing loci was observed in the *C. cereale* population, with 11 of the primer pairs (42%) resulting in no PCR product from one or both of the two clade B isolates, despite the ability of these oligonucleotides to prime amplication from the majority of the clade A populations (Table 5.3). The absence of these 11 microsatellite loci in clade B was consistent with the well-documented diversification of *C. cereale* clade A and B populations (Crouch *et al*, 2005; Crouch *et al*, 2006; Crouch *et al*, 2008a; Crouch *et al*, 2008b; Crouch *et al.*, 2008c; chapters 1-4). The absence of the microsatellites reflected the pattern of absent copies of transposons described in *C. cereale* clade B, relative to clade A (Crouch *et al*, 2008a; Crouch *et al*, 2008b; chapters 2-3). The shared presence of the remaining 15 loci in both *C. cereale* clades A and B supported the close connection of these two major groups as members of a single species in spite of the distinctness of the two groups

(Crouch *et al*, 2005; Crouch *et al*, 2006; Crouch *et al*, 2008a; Crouch *et al*, 2008b; chapters 1,2,4).

A relatively high percentage of loci -- 23% -- were unexpectedly absent in isolate CA-1714, a member of the same population as template isolate NJ-6340 (positive control) (A10; see Table 5.3). Six of the 26 primer pairs failed for CA-1714, while the NJ-6340 control and the majority of clade A isolates produced an amplicon. Thus, despite shared membership in the same population (Crouch *et al*, 2008a; chapter 4), the two isolates are quite distinct on the molecular level, hinting at a level of diversity unanticipated within a species for which the sexual state has never been observed and recombination has been inferred only secondarily through population genetics (Crouch *et al*, 2006; Crouch *et al.*, 2008a; chapter 1 and 4). Five of the six absent CA-1714 loci corresponded to loci also absent in one or both of the clade B isolates, suggesting the possibility of recombination between CA-1714 and a member of clade B.

Of the 26 microsatellite primers that resulted in PCR amplicons, 22 of the banding patterns were polymorphic when visualized on agarose gels (88%). Microsatellite-containing loci C19, H29, K3 and K60 all produced monomorphic banding patterns in the sampled population; but because these three loci only produced an amplification product in two or three isolates each, the invariability may be a reflection of the small sample of amplicons. Microsatellite H29 was present only in the three turfgrass specialist populations (A7, A8, A10) suggesting a common origin for these populations; there was no discernable pattern of presence/absence in the other three monomorphic loci. It is possible that one or more of these loci might yield polymorphic data if used to screen more extensive samples among populations.

From the 22 remaining microsatellite markers, the size of the PCR products, while consistent with the expected amplicon sizes, were polymorphic between isolates, even as visualized using agarose gels (Fig. 5.2). The amplified DNA generated from the 12 *C. cereale* isolates was variable depending upon the individual microsatellite locus (Table 5.3), with a size range of ~10-100-bp between isolates observed. On average, the size range for the markers was ~34.8-bp. Because the isolates sampled in the present study represent 11 known *C. cereale* populations, these data are a strong indicator that the microsatellite markers developed in this study may effectively (a) discriminate between the major *C. cereale* populations and (b) expand analyses of *C. cereale* to the sub-population and/or individual level.

Pilot level PCR-based screening of *C. falcatum*, *C. sublineolum*, *C. caudatum* and an unnamed species of *Colletotrichum* from bahiagrass (isolates 306299SO, 23857ZT, 305428PD respectively; see chapter 4) using 10 of the microsatellite primer pairs polymorphic in *C. cereale* (5C7, A14, A16, A20, A18, D16, F5, F11, M2, M3, C21) demonstrated that all of these primed amplification of microsatellites in one or more sister *Colletotrichum* species (data not shown), suggesting these markers hold the potential for use in population studies of *C. cereale*'s closest relatives. To test whether the microsatellite loci were reliable molecular markers in species other than *C. cereale*, the 35 sequences were used as queries in nucleotide-based BLAST searches against the NCBI Trace Archive of the *C. graminicola* whole genome shotgun sequence (searches performed 17-March-2008; e-value cutoff 1e-20). *C. graminicola* is one of *C. cereale*'s closest relatives (Crouch *et al*, 2008; chapter 4); for this reason, along with the availability of the genome sequence, *C. graminicola* is ideal for these comparisons. Thirteen of the 35 *C. cereale* sequences (37%) corresponded to sequences present in the *C. graminicola* genome (5C7, 7D13, A16, A18,

B8, B9, C15, C21, F5, F10, J6, M2, M3). Examination of the shared sequences revealed that in all but one instance, only the flanking sequences were shared by the two species. Just one of the 13 corresponding *C. graminicola* loci possessed the microsatellite repeat motif characterized from *C. cereale* (Fig. 5.2 A-D). For example, in *C. graminicola*, a gap is present at the location of the *C. cereale* CA_{10} repeat of the 5C7 locus, despite the overall pattern of conservation in the region (Fig. 5.2.A). At the A18 locus in *C. graminicola*, a gap corresponds to most of the *C. cereale* GT_{12} repeat (Fig. 5.2.B). And at the F10 locus, although the *C. cereale* GA_7 repeat is absent in the *C. graminicola* genome, the two species share a CCT repeat at this site: three non-consecutive copies in *C. cereale*, and an uninterrupted CCT 7-mer in *C. graminicola* (Fig 5.2.D). Only the TG repeat sequence from locus F5 was conserved between the two species, with *C. graminicola* possessing 3 repeats, and *C. cereale* with 7 copies (Fig. 5.2.C). Based on these data, I conclude that many of the *C. cereale* microsatellite loci described in this study might be successfully amplified from *C. graminicola* and other closely related *Colletotrichum* species from grasses, but their as molecular markers for any species besides *C. cereale* is not recommended. In particular, these markers are unsuitable for interspecific comparisons. At best, the absence of the repeat sequences in *C. graminicola* suggests the loci will be monomorphic and of no practical value for population analysis of species outside of *C. cereale*. At worst, as with the F10 locus where the characterized *C. cereale* repeat motif is absent and a different repeat motif is observed in two species, these markers may generate homoplasic genotypic data inconsistent with evolutionary history when used outside of *C. cereale*. This finding is consistent with the results of analyses performed in the fungal genus *Neurospora*, where interspecific microsatellite homoplasy and high levels of saturation were frequently observed (Dettman and Taylor, 2004).

In conclusion, the 22 microsatellite loci characterized in this work constitute a promising set of polymorphic markers for fine-scale intraspecific research, phylogeographical analyses, dispersal patterns and other population processes that impact the biology and evolution of *C. cereale*.

5.4 References

Brown, J., Hardwick, L,J. and Wright, A.F. 1995. A simple method for rapid isolation of microsatellites from yeast artificial chromosomes. Mol. Cellular Probes **9**: 53-58.

Carleton, K.L,. Streelman, J.T., Lee, B.Y., Garnhart, N., Kidd, M., and Kocher, T.D. 2002. Rapid isolation of CA microsatellites from the tilapia genome. Animal Genetics **33**: 140-4

Crouch, J.A., Clarke, B. B., and Hillman, B. I. 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. J., **10**: 186-195.

Crouch, J.A., Clarke, B.B. and Hillman, B.I. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology **96**: 46-60.

Crouch, J.A., Tredway, L.P., Clarke, B.B. and Hillman, B.I. 2008(a). Ecological specialization drives the evolution of the pathogenic fungus *Colletotrichum cereale* and allied taxa across diverse grass communities. Submitted to Molecular Ecology 01-Jan-2008.

Crouch, J.A., Glasheen, B.M., Giunta, M.A., Clarke, B.B. and Hillman, B.I. 2008(b). The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. Fungal Genetics Biol. 45:190–206.

Crouch, J.A., Glasheen, B.M., Uddin, W., Clarke, B.B., and Hillman, B.I. (2008c) Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles. In press, Crop Science.

Dettman, J.R. and Taylor, J.W. 2004. Mutation and evolution of microsatellite loci in *Neurospora*. Genetics **168**:1231-1248.

Kõressaar, T., Remm, M. 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics **23**:1289-91.

Mann, R.L., and A,J. Newell. 2005. A survey to determine the incidence and severity of pests and diseases on golf course putting greens in England, Ireland, Scotland, and Wales. *Intl. Turfgrass Soc. Res. J.* **10**:224-229.

van Putten, W.F., Biere A., van Damme, J.M. 2005. Host-related genetic differentiation in the anther smut fungus *Microbotryum violaceum* in sympatric, parapatric and

allopatric populations of two host species *Silene latifolia* and *S. dioica*. J. Evol. Biol. **18**: 203-212.

Raboin, L.M., Selvi, A., Oliveira, K.M., Paulet F., Calatayud, C., Zapater, M.F., Brottier, P, Luzaran, R., Garsmeur, O., Carlier, J., D'Hont, A. 2007. Evidence for the dispersal of a unique lineage from Asia to America and Africa in the sugarcane fungal pathogen *Ustilago scitaminea*. Fungal Genetics Biol. **44**:64-76.

Rivas GG, Zapater MF, Abadie C, Carlier J. 2004. Founder effects and stochastic dispersal at the continental scale of the fungal pathogen of bananas *Mycosphaerella fijiensis*. Mol Ecol.**13**:471-82.

Schuelke, M. 2000. An economic method for the fluorescent labelling of PCR fragments. Nature Biotechnol. **18**: 233-234.

Sexton A.C., Whitten A.R., Howlett B.J. 2006. Population structure of *Sclerotinia sclerotiorum* in an Australian canola field at flowering and stem-infection stages of the disease cycle. Genome **49**:1408-15.

Smiley R.W., Dernoeden P.H., Clarke B.B. 2005. *Compendium of Turfgrass Diseases*, 3rd edn. APS Press, St. Paul.

Stukenbrock EH, Banke S, McDonald BA. 2006. Global migration patterns in the fungal wheat pathogen *Phaeosphaeria nodorum*. Mol Ecol.**15**:2895-904.

Table 5.1.

Microsatellite loci identified from the enriched genomic DNA of *Colletotrichum cereale* isolate NJ-6340 and PCR primer sequences.

Locus	Motif	Primer Sequence	Size (bp)	Percentage of 12 isolates producing amplicon
5C7	$(CA)_{10}$	F-CCGCGACAAACTACGGTTTA R-CTGTCGTCCGACTACCGATT	176	100%
7D13	$(CA)_{26}$	F-AATTGAGGGGTCTTTCTCACC R-GAATGGTGAATGCGTTGTTG	140	100%
8K10	$(TG)_{11}$	F-CCGTCATGACAAACCTGCT R-TCGCTATCGACCTATCTCTTCC	284	100%
A7	$(CA)_{15}$	F-AGGACGAGGATGACATGGAG R-CGGTTTCTTCAAGGCAAGAG	203	0%
A11	$(GA)_{17}$	F-ATCCCTAACCCAGCGTTTCT R-CTGGCATGGCAATCTTGG	163	83%
A14	$(GT)_{17}$	F-GCAGCCTCATCGTAATCACA R-ATGCACATCTCGCATCACAT	183	75%
A16	$(CA)_{12}$	F-TGAGGTAAGGTACGGCTGGT R-CCAAGCCTCTTCCTCTGTTG	198	92%
A18	$(GT)_{12}..(GA)_{25}$	F-GGTGCGTCTTCCCAAAGG R-GAACGGTCAGCAAACTGG	300	67%
A19	$(CT)_{13}(CA)(CT)_3$	F-GCGACCGATTATTGACCATT R-GGGACATGGTGTGCAAAGTA	159	100%
A20	$(CA)_{12}$	F-ATCCCGATCTCCGGTATCTC R-TTGACAGCATTCAAGGCAAG	183	83%
A22	$(CA)_{13}$	F- CGTAACCGGTCCATGTTCCT R- GGGATGTTGCTGCTCCTAGA	219	92%
B4	$(GT)_{21}$	F-CCTTGGGCTGGAAATGGT R-GAATTCACTAGTGATTATAGAATAT	150	0%
B8	$(GT)_9$	F-TCAGCTTTAGCTGCCCTTTC R-CTACTGACACCGGACACCAC	155	0%
B9	$(CA)_{11}$	F-CTCAGCACATCCAGCCAGT R-GCGGTGATGGTGTTGTTG	261	92%
B17	$(GT)_{12}$	F-GAAGCTTGGGATCAGACTGC R-AATACCATCCATCATTCAACATC	149	0%
C15	$(GT)_{16}$	F-GGGGACAAGGGTAAGAGGAG R-CCGCGGGAATTCGATTATAG	232	0%
C19	$(CT)_{15}(GT)_{16}$	F-GACTCACATTCCGAGCCATT R-CCGCGGGAATTCGATT	237	33%
C21	$(GT)_{13}$	F-CTGGCCTCAACCTGAAGAAG R-AAGCTGCGGAGATAAGTTG	232	100%
D2	$(GT)_{16}$	F-TGCACGAGTCGGTCTAAGTG R-CCGCGGGAATTCGAT	138	0%
D10	$(GT)_{10}..(CTT)_5..(TACC)_4$	F-CAGAAGGGGAGAGTACGGATA R-GCCACCAGTCAGTCAGTCG	250	83%

Table 5.1, continued.

Microsatellite loci identified from the enriched genomic DNA of *Colletotrichum cereale* isolate NJ-6340 and PCR primer sequences.

Locus	Motif	Primer Sequence	Size (bp)	Percentage of 12 isolates producing amplicon
D16	$(GT)_{21}$	F-TACGCTCGTCTGCCTCTACC R-CAGATGGCCCCTGTATCAAT	114	100%
F5	$(GT)_{9}$	F-GGCCATGCATCTTTCAGTC R-TCACAAACAGACGGGGTTC	150	92%
F10	$(GA)_{7}$	F-GGAATACCTTGCGCATCACT R-GCTTGGTGCCAAGTTCAGTA	100	0%
F11	$(GT)_{9}..(GT)_{5}$	F-ACAGGACAACGGGACAAGAG R-TCAAACAATCATCCTCACAGC	231	100%
H29	$(GT)_{15}$	F-CCGGCCCCAACCATAC R-GTTCCGGGATTATGATGGTG	100	33%
J6	$(GT)_{20}$	F-CCGAGAAGCTTCAGTCTTGG R-AAGCTTGGGATCGAGGTTG	204	100%
J14	$(CA)_{10}$	F-CCGTTACTGCGTACGGATACT R-TATGATGCTTCCGAGGGAGA	242	0%
J70	$(GA)_{13}(CA)_{10}$	F-CGTGACGATGGGACTGGAG R-GCTTGGCATCTTTCAGTCG	276	92%
K3	$(GT)_{14}..(GA)_{5}..$ $(GA)_{3}..(GA)_{16}$	F-GTGTAGTGAGGAGGGGAACG R-GAATTCACTAGTGATTATAGAATA	241	42%
K41	$(TAGG)_{6}$ $(GT)_{7}(CT)_{3}$	F-TGTAAAACGACGAG R-GGGATGTTTGCTGCT	175	83%
K60	$(GA)_{32}$	F- CCGCAAGGGGGTTTGTAA R-ACTGGGCTGACCGTCCTT	148	17%
M2	$(GT)_{15}..(GA)_{23}$	F-TGTAAAACGACGAG R-TATTGCAGCAAGCG	300	83%
M3	$(GA)_{21}$	F-GCAAGGTTGATGGACTCACC R-TATTGCAGCAAGCGATTCAG	300	75%
M8	$(CA)_{3}..(CA)_{8}..$ $(GA)_{10}$	F-GACAAGGGCAGCAAGAAGAC R-ACATCTCAGGCGAGGGTTT	297	100%
M10	$(GA)_{18}$	F-AGAAACTTGGCCGTGAAGTC R-CCGTTTTTCCAGCCTCTTCT	165	0%

Table 5.2.

Colletotrichum cereale isolates screened by PCR using the microsatellite primer pairs.

Population	**Isolate**	**Host**
A1	24049AS	*Avena sativa*
A2	305377AE	*Arrhenatherum elatius*
A3	KS-TA4F4	*Triticum aestivum*
A4	305429PF	*Polypogon fugax*
A5	305076	*Avena sativa*
A6	IL-CI7.3D	*Calamagrostis inexpansa*
A7	NJ-8626	*Poa annua*
A8	68188LG	"lawn grass"
A9	NJ-HF2B	*Agrostos stolonifera*
A10	CA-1714	*Poa annua*
B	NJ-DG2A25	*Dactylis glomerata*
B	NJ-CA1C1	*Calamgraostis acutifolia*

Table 5.3.

PCR amplification products generated from microsatellite primers at 55°C annealing temperature (absence/presence, designated by +/–). The last column indicates the approximate range of amplicon sizes, if any. Positive control (+) listed in column one is *C. cereale* isolate NJ-6340, a member of population A10.

μsat	***Colletotrichum cereale* populations**													**Size range, in base pairs (approximate)**
	+	**A1**	**A2**	**A3**	**A4**	**A5**	**A6**	**A7**	**A8**	**A9**	**A10**	**B(1)**	**B(2)**	
5C7	+	+	+	+	+	+	+	+	+	+	+	+	+	50
7D13	+	+	+	+	+	+	+	+	+	+	+	+	+	50
8K10	+	+	+	+	+	+	+	+	+	+	+	+	+	15
A11	+	+	+	+	+	+	+	+	+	+	+	–	–	35
A14	+	+	+	+	+	+	+	+	+	+	–	–	–	30
A16	+	+	+	+	+	–	+	+	+	+	+	+	+	70
A18	+	+	+	+	+	–	+	+	+	+	–	–	–	30
A19	+	+	+	+	+	+	+	+	+	+	+	+	+	65
A20	+	+	+	+	+	+	+	+	+	+	+	–	–	15
A22	+	+	+	+	+	+	+	+	+	+	+	+	–	35
A7	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B17	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B4	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B8	–	–	–	–	–	–	–	–	–	–	–	–	–	–
B9	+	+	+	+	+	+	+	+	+	+	+	+	–	40
C15	–	–	–	–	–	–	–	–	–	–	–	–	–	–
C19	+	–	–	–	+	–	–	+	–	–	+	–	–	0
C21	+	+	+	+	+	+	+	+	+	+	+	+	+	20
D10	+	+	+	+	+	+	+	+	+	+	+	–	–	15
D16	+	+	+	+	+	+	+	+	+	+	+	+	+	50
D2	–	–	–	–	–	–	–	–	–	–	–	–	–	–
F10	–	–	–	–	–	–	–	–	–	–	–	–	–	–
F11	+	+	+	+	+	+	+	+	+	+	+	+	+	50
F5	+	+	+	+	+	+	+	+	+	+	–	+	+	25
H29	+	–	–	–	–	–	–	+		+	+	–	–	0
J14	–	–	–	–	–	–	–	–	–	–	–	–	–	–
J6	+	+	+	+	+	+	+	+	+	+	+	+	+	50
J70	+	+	+	+	+	+	+	+	+	+	–	+	+	10
K3	+	–	+	+	+	+	–	+	+	–	–	–	–	0
K41	+	+	+	+	+	+	+	+	–	+	+	–	–	25
K60	–	–	+	+	–	–	–	–	–	–	–	–	–	0
M10	–	–	–	–	–	–	–	–	–	–	–	–	–	–
M2	+	+	+	+	+	+	+	+	+	+	–	+	–	100
M3	+	+	+	+	+	+	+	+	+	+	–	–	–	25
M8	+	+	+	+	+	+	+	+	+	+	+	+	+	100

5C7 (CA_{10})

GTGACCTTGCCCACACGAGCCAAGGCCTCTACCCCTCCTGCCGCGACAAACTACGGTTT
ACCGCCCGCCAGGCCTCTCTATCGATAAACCACACACACACACACACAAACGCCAT
AGACAAATACCCCTCTGTTCTTCCTCTCCACCCTTCAGGCGTCAAGTTTCACTACGATG
CGTTCAAGGCCTTTTTTTCAATCGGTAGTCGGACGACAGCACAAGTATCATGCTTACTC
GAGTCAGCCGCGTGTATTCTAGGCTTCCCTCTCACGGNCTGGGCTAGACCGACACAGTA
TACGTGCGTGCGGTGCGGCGCTCGGCTCGGTGCGGATAACGGCTATCCGCGCTGATTAC
CACMACGAGGGGCGGCAACCCGGTCGAACCCTGGGCGACNGCGGAAGTGCATGCATGCA
CGGCGCCAGCCCACGC

7D13 (CA_{26})

CTGCTGGCCATGTGTTCCCCACACCCACCCCGTCTGGAGTGAGAAAATTGAGGGGTCTT
TCTCACCGAGCAGGCTTCCTCACACACACACACACACACACACACACACACACACACAC
ACACACACACACAGTTACACTACTGGCTGGATTTAGTCTATGTGCTTCAACAACGCATT
CACCATTCTTGCCATGTCAACAACTGGCNACTCGCATTCAAATGGTCCACCCACGTGGT
GCTCTCGCCGCAGCCCAACCTCCCTTCCCCAGTCTCTCTTTCCGTGCTAGGCTCCATGC
CAGGTGTCCTGCCGCGCCCCCCCTCTGTGGGATGGCCTCCGCGATAAGGGGACCCAAGC
CCTGGTTCGGACC

8K10 (TG_{11})

GGGATCCCATAAGCCTGCGTGGTTTTNCCCAAGGCGCGTTTCCGACGGCAGGCCGTCAT
GACAAACCTGCTTTTCTCACTCTTTGACCTTTTCCTCCCTTCCTTTTGTCTTCTTTTTC
CGTCATTTGTTTTTAAACTCTTTTTCTACGGAGCTATCCCAATCTGATTCAACCTHTGG
TCGGGCGCGGTGTTGCGTGTCCGTCATTCTGTTGAGCTCATCGACCCGTCCTTGATNTT
TTTCAAGGCAAGGCGGTCCACGCGAGTTCGCGGTATCCGGGCATGTGTGTGTGTGTGTG
TGTGTGCGTGTTTGTAGAGGGAAGAGATAGGTCGATAGCGATTATTCCAGAGTCGAGAT
ATTCACTTTTTTTTTTTTGGTTTCCGCGTCACAAACTTCTGAGATACCGGAGATCCCAA
GCTTCCCGGGTACCGCAATCGAATTCCCGCGGCCGCCATGG

A7 (CA_{15})

ATCGCAGTGAAGGCAAACAGAAACAGATTAAGTGCGGTGGCTATTGCYGTTTGATGGGG
TGTACGGATAGGGGCAAACGGAAACAGGACGAGGATGACATGGAGACGAAGAGCAAGTA
GTGAGAGCGGTGTGGTTGGAAAAAAGGAAGGATAAATGGAGAGTAGACGTAGCTTTCCT
TCTCTGCGAGAGAACTAACTGCAGGTAAGGATAGAATGGAAGGCACGGCCAGTAACACA
CACACACACACACACACACACACACATACGGCTCTTGCCTTGAAGAAACCGCCCACTTG
GAACCGCCACCTCACGGAAGGTTCAACGGGGAGGTGCGTGGCGCCCCAGCCTCCGGATG
GAAAGATCCCAAGCTTCCCGGGTACCGCAATCACTAGTGAATTC

Figure 5.1.1. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primers sites in gray.

A11 (GA_{12})

ATGCGTGAACTCCGACATCCCTAACCCAGCGTTTCTCCCNCAGCTTCCACCGGAGNAGA
AATCAGCCTCAGCTCGCCTCGCTTCCCCAGGCAGAAACTTGGCCGGGAAGYCGCCGAGA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTCTGCDGSTGCCAAGATTGCCATGCC
AGCRCGAATGGNGSCAAAAAAAGGCCCGTCMTCYGATCTTTCT

A14 (GT_{17})

CATCTCTTGTCGGGTTCGTCACAACAGGGACAAACGGTATCATCMATGTTGCGATATTC
ACCCCGGTTGCCAGCCAGCGTTGCAGCCTCATCGTAATCACACGGTGACGCACCACGGG
AGAACGGCAGAACGGCAGCTGGGGCCTTTGCCGTTTTGGAAGACTACACAAGCTACCGG
TCCCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGGGAGCGAGCAAAAGATGGT
GGTGTGCGATGTGATGCGAGATGTGCATAAGGTTGTGACGATGGGYCATGTCGACTCCC
AA

A16 (CA_{12})

ATCCGTCGTTGGAGATTCTGGTTTGGGATGAAGAGGGTCAATCTCNCCGAGCGGCAACC
ACCGGCACGATTGGCAAGGAAGACGGGCTCAAAACGGAGCCAGGAGGGGCCATGGTATG
GCGAGCCAAGCAAAGAAAAAGGGAAAAATGAGGTGCAGGGCTGGGCAAGGCGCGAATGA
GGTAAGGTACGGCTGGTACGTGCGTCCTGCTGTGCTTCCCTGTCACAGGCCATGTACTC
AACCACTAAGAAAAGGGACTCTACACCGACCGACTGGAGTCAGATTGCAGTCAGAGGGA
GCGCGCGCGCACACACACACACACACACACACATACGCACTCTCTATCTCACGCTGACA
ACAGAGGAAGAGGCTTGGACCATTACGAGAGCGCCAGGTGTGTATCTGGCTTGTGCCTT
GTACGGACCATTGGTAGGCTAATGTACGCCTGGGTTATAGGCGGGAGCCAAGACAATGG
GTTCTGTGCTCGTTATTTACTCGCAGGAAGAGGATGGGTGGAACATGGTCAGTCTCTCG
GTCAAACCAGACATCGGAACAAAGTCATGACGAGTGTGCGCGA

A18 ($GT_{4...}$ GT_{15} ...GA_{25})

CCCCGACCAACATGGGGAATGCGGCATTGATTGGTTCTTTTTCCYTGGTGCGTCTTCCC
AAAGGGGGGTGTATATGTGTGTGTATATATGTGTGTGTGTGCGACGCCACCTAAGACAA
CATSAGCAGAGAATTTGCAATCCTGGCCCGTCTCTGCTGTCGTCAGTATGCCCGTCTGN
CTCAAAACACCTCATGCACGTCAACGGCACTACCTGCAGCAGGCTGGACTAATCATACT
GCAGGAAGAGAGAAAGAGAGAGAGAGAGAGAGAGAGAGGGAGAGAGAGAGAGAGGG
AGGGGGGNCCCCAAGAAGGGCGTGCCNAAGCCGCCAGTTTGCTGACCGTTCCCGCGTAC
GGGACGNTAGSAGGCTACACGAGCCCGCTTCACAATCAAGACAGGACCCT

Figure 5.1.2. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primer sites in gray.

A19 (CT_{13} CT CT_3)

CCCGTCAAGCGGTGTCATCATACCTTGTCTTACCCGAGTTAGAAAGGGAGCGCGAATGC
CGAGCCTGCCTACGGAGGTACTCTCTGGTTCCCTGTCCCCCCCCCTACCTCTCCCCTTT
CGCTTCGCATAGTACCCACTGAATGCCGGTCACCGGCGACCGATTATTGACCATTGAAA
AAAATTCCGTGTGACCCGGCGACGGCCCCGCATAACTTTCCACACCACGGGTGCCTTTT
CTCTCTCTCTCTCTCTCTCTCTCTCTCACTCTCTACCTCTTCCCCGGAAGACTCTGTAC
TTTGCACACCATGTCCCATGATCCCAAGCTTCCCGGGTACCGCANTCACTAGTGAATTC
GCGG

A20 (CA_{12})

CGGGAAGCTTGGATCCCGATCTCCGGTATCTCAGAAGTCTGTGACGCGGAAACCAAAAA
AAAAAAGTGAATATCTCGACTCTGGAATAATCGCTATCGACCTATCTCTTCCCTCTACA
AACACGCACACACACACACACACACATGCCCGGATACCGCGAACTCGCGTGGACCGCCT
TGCCTTGAATGCTGTCAAGGACGGGTCGATGAGCTCAACAGAATGACGGACACGCAACA
CCGCGCCCGACCAAAGGTTGAATCAGATTGGGATAGCTCCGTAGAAAAAGAGTTTAAAA
ACAAATGACGGAAAAAGAAGACAAAAGGAAGGGAGGAAAAGGTCAAAGAGTGAGAAAAG
CAGGTTTGTCATGACGGCCTGCCGTCGGAAACGCGCCTTTTGCGAAACCACGCAGGCTT
ATGGGATCCCAAGCTTCCCGGGTACCGCAATCACTAGTGAATTC

A22 (CA_{13})

CGAAACGAGCTCCGGTGGGATGGTCAGGAGGCCCGGTTGCYGTGATCCCCGTAACCGGT
CCATGTTCCTACCACACACACACACACACACACACACAGGTCCTGCAACACTCACGCAA
CGTGGGCTGCTTTCGAAGTTCTGAAGAGGGGGGGAGGAAGTCGGGGAAAAAGACGCAGG
TTCGCGAGGGGAGAGGGGTGTCAGCAATCATGCGCGTGGATGTAAAGAGGTCAGCTTAT
GTAAGATATACATCTAGGAGCAGCAACATCCCGTTCCAATCGCGAATATGATATGTGAA
GCCTCCACGGGCTGAGCAGCAGATAGCTGATAATGCATGATCCCAAGCTTCCCGGGTAC
CGCAATCACTAGTGAATTC

B4 (GT_{21})

AGGGAGCAACGGCCTTGGGCTGGAAATGGTGTAGTGAGGAGGGGAACGGTAAACGAGGA
GGGGAGGGGGAGGCAGGCCAGGTATGCTACTCTGCGTCTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGATATTCTATAATCACTAGTGAATTC

B8 (GT_9)

ATCAGCAGTGTGTACGTAGTCAGCTTTAGCTGCCCTTTCGGCTKTGCCGCCGGGCGTAA
TAAATTGTGTGTGTGTGTGTGTGTACAGTACAGCCTCCAATGAGCACACTGGATGCTGC
TTGAAAAAGCAGTAGGACGATGTGGCAGACCTTCTCGTGGTGTCCGGTGTCAGTAGGAG
TGGTGGGTGYGGTGCGGTGCGTCGCGCGCGGNCRANAACCGCGGCTGGACGGTGCTGTG
AA

Figure 5.1.3. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primer sites in gray.

B9 (CA_{11})

CGGTACCCGGGAAGCTTGGGATCTCCCTTGTANCTGGACCKKGMCCACCGGTAATGTAC
TGAGTATTCGTCAAAGGCGTATGCGTACACTTTAACAGGTATACTACGTAGTTTACTCC
GCATTGGTCAACTCTTCTACTTGGTGGAACGCGATGTTTACGAGAGTATAAACTATCCG
GNGGGTGCTCTACTCTCAGCACATCCAGCCAGTCACTCAGCATCAACAAACCATCACGT
TCTTTACACACACATTGCCAACCAGCAACTACTTGTTCAAAACCACACTCGCTTCCAAA
CACACACACACACACACACACACGCACACTCGCTTTCCAAAATGCAGACCGCCCTCCTC
TTCCTCGCCACCCTCCTCGCCGGCGTCAACGCGTGCGACAGATACGAGAAGTGCCACTG
CACCAACGCCAACGGCTCCGCCAACAACACCATCACCGCGACCGTGTGCAACAACGCCA
ACCAGTTCGCGCCCGATGCCCCCGGCTGGCCCGTCTACTACCTGGACGAGAATGATGTN
CACCAAGGGGCATCGTCAACACGAAAACCNACCTTCGCCCCTCCA

B17 (GT_{12})

TGCAGGAACCCGGGAAGCTTGGGATCAGACTGCTTTTGGTAGTAACATTCCCTGGAATT
AGTCTCCCAAAACTTTGTCGCTTCGGGTTGCCAAAGTGTGTGTGTGTGTGTGTGTGTGT
GCGCATGTGTATCCAAATCAAGATGTTGAATGATGGATGGTATTMTGCTTCATTAAGGT
AACTAGTGTACTCTGTGTACATACGGGGTCGCGTCCGCCTGGTTCGTGGCGACTAGGTT
TGTACAAGTAGTAGTGTAGTTTTTTYCCAGCCGTCGCTACGATTGTCGACGGGGGGAGC
AGGTTGTGCTGTTGTGTCAATCTATCTCTCTTCCTCTCTCCGTCTTCCTTCTAAACACG
ACTGTGACATTTTATCGACCCGATTCGTCCCTTTTCCCCCCATTTTTAACAAAACGGAT
TATACTTCCATACCCGTTTTACTTTCCGGGGAACACGGATTCCAAGCTTCCCGGGTACC
GGCA

C15 (GT_{16})

GGGATCATGGATGGTGGTACGCCGCCTCATTTGCTCATTGGAGGGGGATGGGAAACTAA
AAACAAAAATTGCAAAGAGGCGAATCCTGGGAACCCAGGCCCTAGGGGACAAGGGTAAG
AGGAGAGACGACTCCCTTTCCAGGGTGGTTCTTCTCTCTCTCCCCGTATACCCGTATCC
GAACTTTCTTGAATGCTAACACACGCGTCTCCCCCCCGCAACCCCTTCTATGCGTCTTG
GATTTAGACATTCCGAGGAGGATTACTATTGGGACACCCTGGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGATATTCTATAATCGAATTCCCGCGGCC

C19 (CT_{15} GT_{16})

TACCCGGGAAGCTTGGGATCCGGTCGGGCGTAGCTGATCGACTCACATTCCGAGCCATT
GATCGTCGTCAAACCCTACGGAGGTTTGGATGTGCTGGCCTGACGATGCAATGCTACCT
ACATCTTTCCCGGACGAGCAACTCTCGCGAGGGTCCAACATGCCCGTTTGCCACTTTAC
CCTGTTGGTGCCCGAGGCCTTCTCTCTCTCTCTCTCTCTCTTTCTCTCTCTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGTGTAATCGAATTCCCGCGGCC

Figure 5.1.4. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primer sites in gray.

C21 (GT_{13})

GTACCCGGGAAGCTTGGGATCGTCAAGTCATTCCTGCCNTTTCCGCCTGCCTGGGCTGC
TGGAACAGGCGGTATGCGCGTGCTTGACTGGCTGGCCTCAACCTGAAGAAGACTGTCAT
CTGCGCCGATCTCCCGCCCCGGAGGAGGGGAGGAAACTAGTGACGCCAGATGGACAGCA
GTCTACTGCAAGTCCAGGTGCGCATCGAACCTTTTGGTTGGTCGGGCCCAGCTTTCACA
TCAATGTGTGTGTGTGTGTGTGTGTGTGTGCGTATATGTATGTATGGGTGTATAGAGYA
CGCGCGCCCAACTTATCTCCGCAGCTTGGGGAACTGAGGACCCTYGGAGGAGCTGGTCG
ATCCCCAACCTTCCCGGAATCGAAATTCCC

D2 (GT_{16})

TTCNTCACTAGTGATTGCGGTACCCGGGAAGCTTGGGATCAACCTACAAGCAATGCACG
AGTCGGTCTAAGTGGGTGCTTACCTCTTCTCGAGAGCTTGGGGGAGTTCAAATGCCGCC
ATGTTTGACGGTGGCCGAAATGTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAA
TCGAATTCCCGCGGCC

D10 (GT_{10} ...$TACC_4$)

AAGGTACCCGGGAAGCTTGGGATCCGGAAGGGTGATTGGTCTGACCAGATTGGGACTCG
ACATGCCTTACCTTTGGACGGAGTGTCTGGCTGTTCTCATCATCGTTTCATGGGATAAG
CGACATGCGATGACGGAGCGAACTCGGCCACAGAAGGGGAGAGTACGGATACTCTGTAG
ATACTTGGGCAGAAGTGTAAGCGTGTAGGTGTGTGTTTGTGTTTGTTTGTAGCTGACGG
GACGGGGAGGGGCCGCCCAGCATGTCACTCACTGCCTCTTTCCTTACTTTCTTCCTTCC
TCCTTTGGGCGGTATCATATCAATTCTTGAACTCTACCTACCTACCTACCCTGTACTGT
ACGTGTAAACGACTGCGGGCGTAGTCGACTGACTGACTGGTGGCCCCTTTTTCACAAGC
TGCCAGGGCGTTTTTGACGGGTGACTGGCGGATGATTTTGCCCCTAGTGGCAGCATTAS
TCTTGGGGCGCAAAAGAAACTTGCGCGAGGATGAACGGCGSGAAGACTNATGACAGCGA
TAGTTAAACCNAGATAAAAAGAAGAATAACCCTGCTGGTCACCTCCCCGCACGCCCCAA
ATCCCCAGAACG

D16 (GT_{21})

AACCNGGGAAGCTTGGGATCCTGTTAACCCCCCCTGGTTTGGCCSGCTGAGCATTTGGT
CGCGTACCCCTGCTCATGCAACCGCTCTCGGAGAGGGTACCTCCAGCCCCCAGCTACGC
TCGTCTGCCTCTACCTACCTACAGGTACATACTTACAGTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGCAGCGACCTACATTGATACAGGGGCCATCTGGCGGAAGAC
CAACGCAG

Figure 5.1.5. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primer sites in gray.

F5 (GT_9)

CTTGGGATCACGAATATCTACGAGGAAGCTGGGCTTATTAGTCTCGGTCGGGAAGAAAA
CAAGATGAAAACACCCTTTTTTTTTTCTCGAGAGGGCAGAATGGGTTGAAAACACATGG
CAAAAAAAGGGCCATGCATCTTTCAGTCGTTTTGTGTGTTGTGTGTGTGTGTTAGCGCG
GAACAACGCTAGATACTTTCAATAGGTTTTTTCTTTCCCTCGAACGCTTTCTTCCCTCC
TCCCGAGTGTTTACCCGGTGGTGAACCCCGTCTGTTTGTGAGTGTTTACGTGTTTCCCG
TTACCGGGGTGGTGCCGGCCCGCGTGAGGAGAYGKCGATCCCAAGCTTCCCGGGTACCG
CAATCACTAGTGAATTC

F10 (GA_7)

AACCCGGGAAGCTTGGGACCTGGAGCCGGCCGCATGTGGTTCGCCANCCACGATAAGAC
AGCTTGTCAGTCTCGGGCGCAACCGGAACAGGCGCACTTCGCACCGACGGTTGACCCTC
CTACCTCCCTCCTGGTTTCGGTGACTACGATCACGGCCAACATGGTACCAGGACACGGA
ATACCTTGCGCATCACTAAGATGTTCGTTCGCGTTGAAGGAGAGAGAGAGAGAGCGCGC
CGTTCTCTCGCCGTTTTNTACTGAACTTGGCACCAAGCGCAGGTCCATGACCGCATCAC
CAGTCCCTCCATCGCTTACGCGCCTGATCGATCCCAAGCTTCCCGGGTACCGCAATCRA

F11 (GT_9)

AACCCGGGAAGCTTGGGATCGGCATTGTCGCGGCCCGAANTTCCCCGGCAAATAGGCAG
CAGGCAATCTTACGGGACAACAGGACAACGGGACAAGAGCTGGGCTGTGTGTGTGTGTG
TGTGTGGCCAATCAAAGTGAGACCATATGACGGGCTGATGAACTTGGACGACAAGGTTC
TGGTGCTTCAAGACGGGAGTGTAVTGTTATGTGTTGAGTTTTTTTTGAACCCGACCCGT
TGGATGAATGATTNTTTTANTGAGTGGACGANAGAGCGTGTGTGTGTCCCGCGCTGTGA
GGATGATTGTTTGATTTGASAGT

H29 (GT_{15})

GTCCCGGCCCCAACCATACGCGTGTGTGTGTGTGTGTGTGTGTSTGTGTGTCCGTCTCG
TCCACCCCGAGCCTCCCACTCGGCCACCATCATAATCCCGGAACCACGAAAACCAAAAC
AAGAAATCAAGACCTCGCGTTGCCATGGGGCCTCTCCTTCTTACGTCAGTGACATGTCT
GGATCCCAAGCTTCCCGGGTACCGCAATCGAATTCCCGCGGCC

J6 (GT_{20})

ATCCCCGTTTACACTATGGTTGGGTTTGCCAGCCTGAATCTGCCCTTCGTCTGAATGGA
AACGACATTGAAACTGGGTACTCTGGGGGCTTGGCCGAGATGCAGAGGGCGAGCATCAT
CTCGGATGGAGACGGGCTGAACCTTGCCCGAGAAGCTTCAGTCTTGGGACGGTTGCGTT
GCAAGGCAGGAGCTGCTGGTAGTGCTGCTCGTCGGGCCGCAGACAAGGGAAAAAATTTG
AAATAATTGTGGATACGCTGACTGTTTATGTTTGCGCGCGTGTGTGTGTGTGTGTGTGT
GTGTGTGTGTGTGTGTGTGTGCGCCGAGGGGTGGTCAACCTCGATCCCAAGCTTCCCGG
GTA

Figure 5.1.6. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primers sites in gray.

J14 (CA_{10})

GGAGTCGACCGGCTCGTTCGCCDGGGGGGTTGTTTNANCCGTTACTGCGTACGGATACT
GGTATGTACAATAATAGTCTAACAACGTGGAATAACGTGTACATGGTATGTAGAAACCA
ACAAGGCAGTTGTGCAGTGTAAGCCAAGAGCATGAACCGAGACGGGCCGGGTGGGAAGT
ACGGATACGTGGCGGTAAGACAGGGCTTCATCCCCCTTTGCGTCATGGTTCTTCACACA
CACACACACACACACGGGHATCATTCTCCCTCGGAAGCATCATACACCAA

J70 (GA_{13} CA_{10})

GATCTCCCTGCCACCGTCCGAATCGTGACGATGGGACTGGAGAGACCGAGAGAGAGAGA
GAGAGAGAGCGCAGGAGCGGCATAAACGTGGAGATCGCCATCTCCTCACGCGGGCCGGC
ACCACCCCGGTAACGGGAAACACGTAAACACTCACAAACAGACGGGGTTCACCACCGGG
TAAACACTCGGGAGGAGGGAAGAAAGCGTTCGAGGGAAAGAAAAAACCTATTGAAAGTA
TCTAGCGTTGTTCTTTTNCTAACACACACACACAACACACAAACGACTGAAAGATGCC
AAGCYTCCCGGGTACCGCAATCACT

K3 (GT_{14} ...GA_3 GT_{16})

AGGGAGCAACGGCCTTGGGCTGGAAATGGTGTAGTGAGGAGGGGAACGGTAAACGAGGA
GGGGAGGGGGAGGCAGGCCAGGTATGCTACTCTGCGTCTGTGTGTGTGTGTGTGTGTGT
GTGTGTGTCCAGGTACCTACCTACCTTGGCTGAAGAGCATACGGAGTATTCTGTACGGA
TACCGGTAGAGAGAAAGCGAGAGTGAGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGG
TGTGTGTGATATTCTATAATCACTAGTGAATTC

K60 (GA_{32})

GGGCCGACGTCGCATGCTCCCGGCCGCCTTGGCCGCGGGATTGCAATGCAAAGCAGAAA
AACTGCACTGGTCGCGCGCACGACGTGACTGTGATCATCCGCCTGTGGATACATCACCC
GCCCGCAAGGGGGTTTGTAAGGCGAGATGAGATGCTGGATCGGAGAGTGAGAATGAGAG
AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAGTGGACAGAAAAGGACGGTCAGCCCAGTGAGGAGGCCTTGTAGCGCCCAGCATCT
AGGTACCTAGGTACCTAGTAGGTAAGGTAGTAGGTAAGACAGCTNACCCANGTGGTGAG
AANGCGCACCGGTACTACCCAGCAGGGAAGCA

M2 (CT_6 ...CCT_3 GT_{16})

GGGCTTCATCTGGGCATGGTCGCCGTCGCAAAGTTGGAGGGGGGAGGGATGCGGTGTTG
ATGGGCTTGGGGGACTTGGAGAGGAGCCGAGATGGGCAGACGGTACGGATTCACTGACA
AGGCGAACGTATGGCTGTCCGAGTGACCAAGTACGATCCAATGCCACAATACAGTTGAA
CTCTCTCTCTCTCTCCCCCCCTCTCCTCCTCCTCGTGTGTGTGTGTGTGTGTGTGTGTG
TGTGTCTCACACATACTCTTCCCCTTCCAAGGAACACTTCCAGAGAACGCCGAATCGGG
GATGGATCACCACGCCATCGGCCTCACTGCAGATCTGGACTCGATCCGTTCCTTCGCTG
TCTTAGATCCCTTCCAGCCGCGACGCCAGACTTGGACATCAGGACTTAAGCTTATTGGG
TGGGCGACGCCAAAGAAGCGTCCCCTTTTCCGTCCCTTACCACAAGGAGGGGGAGGGGG
GCATGTGAGCCCAAGCTTCCCGGGTACCGCAATCACTAGTG

Figure 5.1.7. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primer sites in gray.

M3 (GA_{21})

CAGCTTGGGATCACTGATGTCCATATAAGCATATTCCTAATGTAAGATACAAGGGGGGT
TTGGCCGTGATACGGGTGGGCATTAAGAACAGAGAAGTGGAGGTGGAGTTCAGAGTCGG
GGCCAGGGACGGGGGAGAGTCTTGGGATAAAGCAAGCCGGGGTGGTGTGGTGCGTGTGT
AAGGCGACGGCATGTGGTGTCGACGTGTGGTAGGTTGTTAGGTGTACTTGGGCAGAATG
GCAAAAACAACGAATGGCTGGGGGCAAGGTTGATGGACTCACCGTCGTCCGAACGTTTT
CCCACAGCGAGCCCAACCGCAGGCGAAGGGCTTCTCTTTGGTGTGCGTCCTTTCGTGTC
TTTGAACATGCTCCACACGCCTAAAATGTAAAAGCAATGTGAGCTAGTAAAAAGGTACA
GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGTAAAGGCCGA
TGTAGGGCAAATGTCAACCAGATTGGGGGTGTGTGTGACTTGAACAGGGGGGGGGGGCA
CAACTTACCTGAATCGCTTGCTGCAATACTTGCACGGCAACGTCTTGGGTTTCTGCTGC
CGGGCCTCTCCATTCTCCCCAGCAGCAGAGGCTCCGTGAGGAGGCATACCTTCTATAGA
CTGCAGAGGGA

M8 (CA_{3} ...CA_{3} GA_{10} TCC_{22})

GTACCCGGGAAGCTTGGGATCACCAGCTACGCAAGGCGAGTWCGCCAAGGACAGCGACA
AGGGCAGCAAGAAGACGAGATCAAACCGCCAAGCTCGCTCTTGAAATCCATCAAACACA
CATACACACACACACACACAGAGAGAGAGAGAGAGAGAGAGGCGCGTTTCGACCTCGCC
CTCAGTCAACCTCTCCTATCGAGACGCCACACACACATCCCCCGTCCGAATAGGCTTTT
CTTCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCTTCTTCTTCTTCTATGAGTGGATTTCCCATCTCGTGAAACCCTCGCCTGAGATGTCT
TCGTCCCATTCTCTCAAACCCGTAAGATCCCAAGCTTCCCGGGTACCGCAATCGAATTC
CCGCGGACGCCATGGCGGCSGGGAGCATGCGACRTTCTGNCCCAATTCCC

M10 (GA_{18})

AGAAACTTGGCCGTGAAGTCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GTCTGCAGCTGCCAAGATTGCCATGCTAGCACGAATGGGGCCAAAAAAAGGCCCGTCAT
CCGATCTTTCTCCACCAACCGGGCGGAAGAAGAGGCTGGAAAAACGGCTTCGAGGCCGG
AACAACGGGCTGCTTCCGTCGTCCCAGAGACNGGGATCCCAAGCTTCCCGGGTACMGCA
ATCGCTAGTGAG

Figure 5.1.8. Sequence of plasmid clones containing microsatellite motifs that were also suitable for PCR amplification. Microsatellite motifs are highlighted in yellow, primer sites in gray.

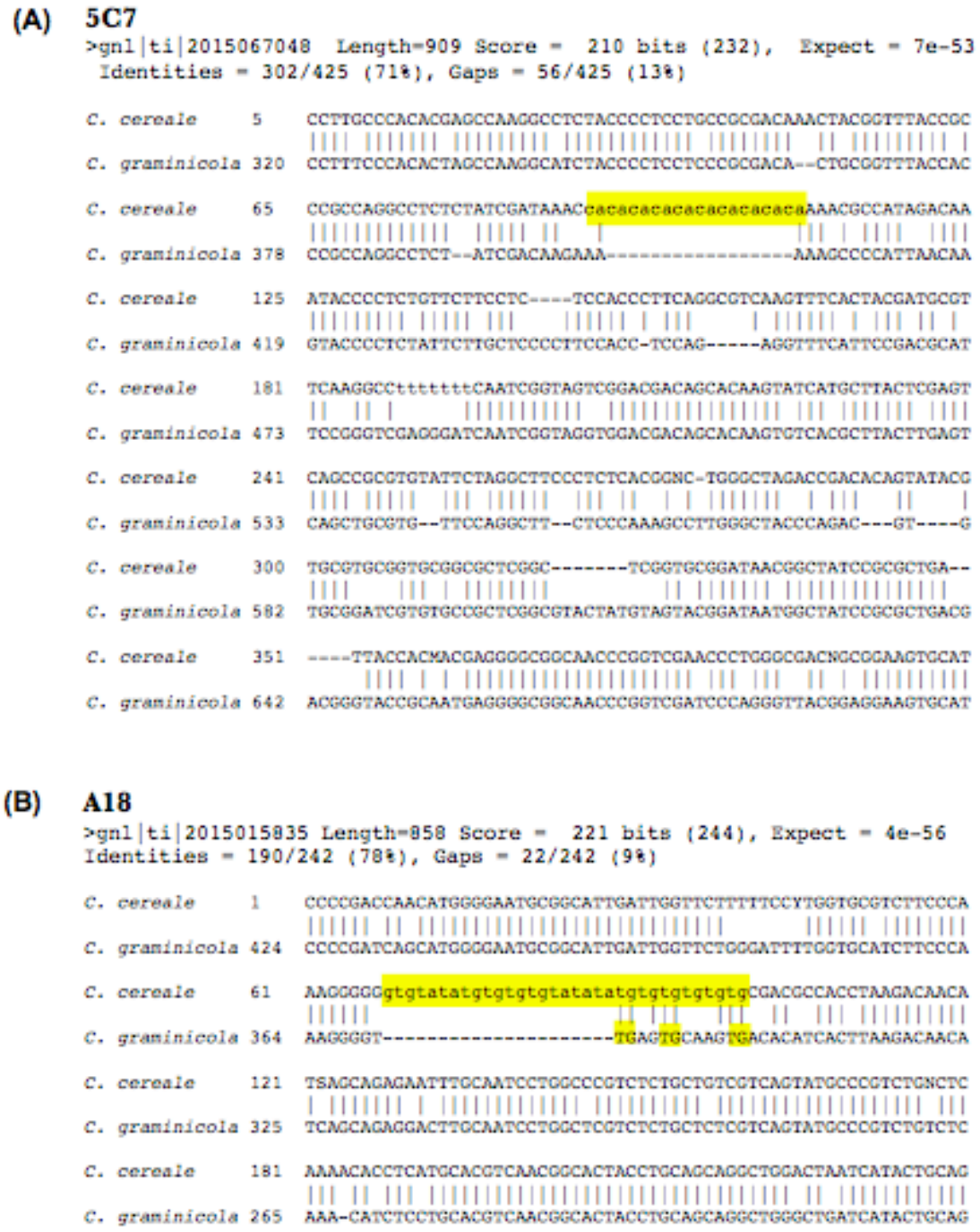


Figure 5.3 (A) and (B)
Representative alignments generated from BLAST searches of the NCBI Trace Archive of the *C. graminicola* whole genome shotgun sequence using *C. cereale* microsatellite-containing loci (A) 5C7 and (B) A18. Although the flanking regions are very well conserved at both loci, the microsatellite motifs (highlighted in yellow) are absent in the *C. graminicola* genome.

Figure 5.3 (C) and (D)
Representative alignments generated from BLAST searches of the NCBI Trace Archive of the *C. graminicola* whole genome shotgun sequence using *C. cereale* microsatellite-containing loci. (C) at F5, three TG repeats are present at the *C. graminicola* relative to the seven in *C. cereale* (highlighted in yellow . (D) At the F10 locus, *C. graminicola* possesses a CCT7 motif very different from the CTT repeated at the same location in *C. cereale* (highlighted orange). The GA7 microsatellite motif from F10 is absent in the *C. graminicola* genome (highlighted in yellow) .

Chapter 6: Systematic analysis of the falcate-spored graminicolous *Colletotrichum* and a description of five new species from warm season grasses

ABSTRACT

Species limits in the fungal genus *Colletotrichum* are traditionally distinguished by appressorial and/or conidial morphology or through host plant association, but both criteria are criticized for their inability to resolve distinct taxa. In previous research, eight novel falcate-spored *Colletotrichum* species were identified from warm season grasses hosts using multilocus molecular phylogenetic analysis. In the present work, formal descriptions and illustrations are provided for five of the new taxa, including *Colletotrichum hanaui* sp. nov., *C. nicholsonii* sp. nov., *C. jacksonii* sp. nov., *C. miscanthi* sp.nov., *C. axonopodi* sp. nov.; and an emended description and illustrations are provided for *C. eleusines*. Comparison of hypophodia appressoria and host association against phylogenetic species boundaries and evolutionary relationships in the graminicolous *Colletotrichum* group demonstrated that while these characters could be useful in combination for the purpose of species diagnosis, erroneous identification is possible, and species boundaries may be underestimated, as exemplified by the polyphyletic taxa *C. falcatum*. Appressoria were identified as subject to convergent evolution and were not predictive of phylogenetic relationships. Despite these limitations, the results of this work establish that appressorial and host range characters could be used to generate informative dichotomous identification keys for *Colletotrichum* species groups where an underlying framework of evolutionary relationships, taxonomic

entities, and nomenclature have been satisfactorily derived from molecular systematic treatments.

6.1 Introduction

Colletotrichum (Corda) is the asexual state of the sexual genus *Glomerella* (Spauld. H. Schrenk), a well-defined taxon within the Sordariomycota, the class of ascomycte fungi that produces asci in perithecial fruiting bodies. For taxonomic purposes, the anamorphic *Colletotrichum* state rather than *Glomerella* is the focus of classification systems as the teleomorph has not been observed for many species, and is rare for numerous others. *Colletotrichum* species are distinguished primarily through their acervular conidiomata and the presence of distinctive thick-walled, melanized sterile hyphae known as setae. The production of appressoria from conidial germ tubes or hyphae is typical for the genus, allowing the fungus to attach to host surfaces before tissue penetration.

Although the integrity of the *Colletotrichum* as a generic entity using morphological features is well substantiated, intrageneric species boundaries are far less certain due to minimal variation (Sutton 1980, 1992; Cannon *et al.*, 2002). Similarly, host range criteria are often considered unsatisfactory for species diagnosis, as many morphologically similar species can be found in association with a wide range of host plants, often overlapping with other morphologically dissimilar *Colletotrichum* species. Because the sexual state is inconsistently observed, and female infertility is a common phenomenon even for those few species with a *Glomerella* teleomorph (Vaillancourt *et al.*, 2000), the application of biological species criteria through mating experiments is not a viable alternative. As a result, multilocus molecular phylogenetics are becoming increasingly important adjuncts

to morphological characters and host association criteria for the definition of *Colletotrichum* species boundaries and intrageneric evolutionary relationships (Crouch *et al.*. 2006; Du *et al.*, 2005; Farr *et al.*., 2006). In the present study, multilocus phylogenetic analysis of the *Apn1*, *Mat1*, rDNA internal transcribed spacer (ITS) and *Sod2* sequences (Crouch *et al.*, 2008a; chapter 4), along with morphological examinations, were undertaken on a set of *Colletotrichum* isolated from grass hosts. Molecular phylogenies were used to determine whether appressorial size and shape and host range criteria are consistent with the evolutionary history of the falcate-spored, grass associated *Colletotrichum* group. In this paper, the formal description of five new falcate-spored *Colletotrichum* species from warm season grasses and the emendment of one species is proposed based on the conclusions of these analyses. To simplify discussion, they are identified here as: *C. hanaui* (from *Digitaria*), *C. nicholsonii* (from *Paspalum*), *C. jacksonii* (from *Echinochloa*), *C. misccanthi* (from *Miscanthus*), *C. axonpodi* (from *Axonopus*) and *C. eleusines* (from *Eleusine*).

6.2 Materials and Methods

6.2.1 Fungal isolates

A total of 55 specimens of falcate-spored *Colletrichum* isolated from grass hosts (the FG, or falcate-spored graminicolous group) were assessed for this study (Table 6.1). Fungi were grown and maintained as previously described (Crouch *et al.*., 2006). In addition, 46 samples of *Colletorichum* sharing the falcate-shared spore morphology with the FG *Colletotrichum* but different host range (the FN group, referring to falcate-spored, non-

graminicolous hosts) were evaluated for the purpose of morphological comparisons (Table 6.2).

6.2.2 Molecular phylogenetic analysis

Phylogenetic analysis of the graminicolous *Colletotrichum* isolates was performed using the four genes previously employed by Crouch *et al.* (2008a; Chapter 4) for *Colletotrichum* FG group analysis from the ITS region, the 5' end of the DNA lyase gene *Apn1* ("W1"), the 3' end of *Apn1*, and the 5' end of the mating type gene *Mat1-2* ("M72"), and the manganese superoxide dismutase *(Sod2*). Multiple sequence alignments were generated using Clustal W (Thompson *et al.*, 1994) on the EBI/EMBL site (http://www.ebi.ac.uk/Tools/clustalw), then adjusted to remove all gap sequences and ambiguously aligned regions. Gaps were recoded as multi-state characters and reintroduced into the dataset. *C. acutaum*, a species of *Colletotrichum* with oval spores, was included as the outgroup taxon (Crouch *et al.*., 2006; chapter 1).

Individual gene sequence alignments and the combined dataset were analyzed using maximum likelihood (ML) in PAUP* (Swofford, 2000). Starting trees for heuristic ML analyses were generated from neighbor joining trees with branch swapping performed using tree-bisection reconnection (TBR), and searches were modeled on the parameters estimated using Modeltest v.3.06 (Posada and Crandall, 1998). Bayesian posterior probabilities supporting the ML phylogenies were estimated using MrBayes v.3.0b4 (Huelsenbeck and Ronquist, 2001), with each gene partitioned according to the evolutionary model best describing the data. The Bayesian analyses were performed in three replicate runs for 20,000,000 generations using one cold and three incrementally

heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC). Trees were sampled every 500 generations to calculate posterior probabilities for each branch in the ML tree, and used to construct a 75% majority rule consensus tree after excluding burnin.

6.2.3 Morphological analysis

For morphological assessments, cultures were grown on potato dextrose agar (PDA; Fisher Scientific, Hampton, NH), in plastic Petri dishes (50 mm diameter, Fisher Scientific, Hampton, NH) under constant light at 26°C for 7-10 d. Cultures were photographed when the colony reached the edge of the petri dish. Conidia were harvested directly from cultures and visualized from wet mount slides. Hyphopodial appressoria production was induced by innoculating 5-cm squares of 15% water agar on microscope slides with small sections of mycelia on all four sides. The inoculated agar squares were overlaid with glass cover slips, then placed inside Petri dishes containing filter paper moistened with 1-ml distilled sterile H_2O. The Petri dishes were sealed with parafilm and incubated at room temperature on a laboratory bench. Formation of appressoria on cover slips took place after 4-7 d; cover slips were then placed onto a drop of lactophenol blue on fresh microscope slides. An Olympus CX40 microscope was used for all measurements with bright field illumination, with measurements recorded from a minimum of 50 conidia and hyphopodial appressoria/isolate; and at least 20 measurements for other structures where possible. Measurements were repeated twice. Photomicrographs were obtained using an Olympus BX41 microscope equipped with a MagnaFire Digital Firewire camera and software (Optronics, Goleta, CA).

6.3 Results and discussion

A molecular phylogeny consistent with substantial subdivision within the FG group was generated from the multilocus molecular dataset (Fig. 6.1). As previously reported (Crouch *et al.*, 2008; chapter 4), the FG group underwent a prominent split during its history that served to separate the cool-season (C3) grass associated taxa (*C. cereale*; FG-C3) from those lineages of *Colletotrichum* that inhabit warm-season (C4) grasses (FG-C4).

FG phylogenetic species were diagnosed through the application of a genealogical concordance method, minimally modified from the protocol communicated by Taylor and colleagues (Dettman *et al.*., 2003; Taylor *et al.*., 2000). Each of the lineages delimited as phylogenetic species in Figure 6.1 met four basic criteria: they were monophyletic, they were supported by standard measures of support (bootstrap and Bayesian posterior probabilities) in the combined dataset analysis, they exhibited no evidence of recombination with individuals outside the group (Crouch *et al.* 2008a; chapter 4), and they were genealogically concordant; i.e., there was no conflict between individual gene trees and/or the tree inferred from the combined dataset. Although eleven distinct *C. cereale* lineages comprise the FG-C3 group, the cohesiveness of this species is uncontested, with substantial evidence of gene flow between the lineages demonstrating that *C. cereale* populations are unified under the auspices of a single phylogenetic species, with the lineages categorized as populations (Crouch *et al.* 2006, 2008a, 2008b, 2008c; chapters 1-4).

In contrast to the FG-C3 group, the FG-C4 group was comprised of twelve well-supported lineages that met the standard for the rank of phylogenetic species (Fig. 6.1; see

Figures 4.4 to 4.7 for individual gene tree topologies). The individual gene geneaologies consistently recovered these thirteen phylogenetic groupings, although the ITS gene tree (Fig. 4.7) contained several taxa that were placed outside of the species boundaries inferred in the combined dataset and the other three gene trees. Of the thirteen phylogenetic species, eight have not previously been identified (Fig 6.1): two species from *Paspalum dilatatum*, one species from *Paspalum notatum*, one species from *Digitaria* spp., one species from *Echinichola esculenta*, one species from *Miscanthus sinensis*, one species from *Axonopus affinis* and one species from *Saccharum officinarum*, which was distinct from *C. falcatum*. Samples of *Colletotrichum* at the New York Botanical Garden Steere Herbarium and the U.S. Department of Agriculture National Herbarium derived from these host plants are identified as *C. graminicola* or *C. cereale* (Figs. 6.11.1 to 6.11.11), but the molecular phylogeny indicates that none of these taxa are conspecific with either of these two species.

Two distinct and highly divergent lineages of *Colletotrichum* were consistently identified from *S. officinarum* (sugarcane). Using morphological and host range criteria, these would have been lumped together as a single species – *C. falcatum* (teleomorph=*Glomerella tucamensis*). In sugarcane crops, *C. falcatum* is responsible for red rot disease, a significant problem for farmers in developing countries. The identification of two discrete species of *Colletotrichum* causing red rot disease, is therefore an important finding that may have implications for disease control measures.

With the exception of *C. caudatum*, each of the FG-C4 lineages was exclusively associated with a single host genus or species, suggesting the hypothesis that host range might serve as an accurate method to identify FG-C4 species. Superimposing host origination data onto the FG-C4 group showed that although host range criteria might

serve as a unique identifier for seven of the thirteen total species, this characteristic will not always lead to the identification of the appropriate lineage (Fig. 6.1). For example, two distinct phylogenetic species are diagnosed using molecular phylogenetics for *Colletotrichum* associated wth *Paspalum dilatatum*, but mere knowledge of host association would not point to which of the two phylogenetic species an isolate might represent. These data demonstrate that while host association data may serve as a valuable adjunct to molecular or morphological data, it is insufficient for the purposes of incontrovertible species identification in the FG *Colletotrichum*, even though species limits are so clearly associated with host plant origin.

Hyphopodial appressoria size and shape, arguably the most successful morphological character used for *Colletotrichum* species diagnoses (Sutton 1965, 1966, 1968, 1980, 1992), was found in overlapping ranges between phylogenetic species of the FG group (Fig.s 6.7.1 to 6.7.4, 6.81. to 6.8.6; Table 6.3). In particular, *C. cereale* possessed a plethora of appressorial shapes ranging from globose to oval to obovate, sometimes in combination with lobes or multiple lobes, and sometimes irregularly shaped (Fig. 6.4.1 to 6.4.4). Uniquely, *C. cereale* strain NJ-HF2B, isolated from *Agrostis stolonifera*, produced long appressorial chains (Fig 6.7.3), a morphology previously described only from *C. dematium* (Sutton, 1980; this study, isolate 680TV). The sizes and shapes of *C. cereale*'s varied appressoria were represented in all of the FG species. Appressoria characteristics from FG-C4 species also overlapped among the taxa. Only *C. graminicola* produced morphologically unique appressoria, and even for this species, morphotypes consistent with those exhibited by the other FG-C4 taxa and *C. cereale* were produced. Comparison of FG appressorial characters with those formed by members of the FN group (Figs. 6.9.1 to 6.9.6) showed that FG appressoria were often quite similar, and sometimes

indistinguishable from those of *C. trifolii*, *C. trichellum* and *C. dematium*. But the complex, irregular and deep-lobed appressorial shape observed from strains of *C. capsici*, *C. circinans* and *C. gloeosporioides* did not resemble any of the appressoria produced by members of the FG group. These data support the concluson that appressorial characters alone cannot provide a satisfactory measure of species boundaries for the FG *Colletotrichum* group.

To determine whether the size and/or shape of hyphopodial appressoria correspond to the FG groups evolutionary history, these characters were plotted across a cladogram illustrating the FG phylogeny. As revealed in Figure 6.10, the FG hyphopodial appressoria are homoplasic across the FG *Colletorichum* group, with divergent species sharing common appressorial morphologies. For example, *C. graminicola* and *C. sublineolum* share a unique appressorial shape relative to the other taxa in the group; however, they are not sister taxa, and their closest relations (*C. nicholsonii* and *C. falcatum*, respectively) possess entirely different morphologies that support the independent acquisition of the irregular appressorial trait in two divergent lineages. Nevertheless, while appressoria characteristics exhibit convergent evolution across the entirety of the FG, there is a pattern of shared evolution in one of the subgroups, where the clade comprised of *C. hanaui*, *C. jacksonii* and *C. axonopodi* hold in common a regular, globose shaped appressorium of a relatively constant size.

Despite the overlap of appressorial characters both among species within and outside of the *Colletotrichum* FG group and the fact that appressorial character are not congruent with the evolutionary history of this organism, some general conclusions about species limits can still be derived from these data. If used in conjunction with host range criteria, appressoria morphology can provide details about the identity of some of the FG-

C4 species with only a small margin of error arising from the association of multiple species with hosts in the *Saccharum* and *Paspalum* genera.

6.4 Taxonomy

Five new species are proposed and described for the isolates of *Colletotrichum* from *Digitaria*, *Echinochloa*, *Paspalum*, *Miscanthus*, and *Axonopus*, which are phylogenetically distinct from isolates of the fungus from *Z. mays* (*C. graminicola*), *Sorghum* spp. (*C. sublineolum*), *S. officinarum* (*C. falcatum*), and the wide host range species *C. cereale* (from C3 physiology grass hosts) and *C. caudatum* (from C4 grasses *Andropogon gerardii*, *Imperata cylindrical*, *Zoysia* spp. and *Bothriochloa bladhii*). An emended description is provided for the species *C. eleusines* (*E. indica*). Although the application of genealogical concordance criteria to define phylogenetic species supported the description of eight novel species rather than five, I have elected not to subdivide the two taxa derived from a single host genus or species as part of this work (*C. falcatum* from sugarcane and *C. nicholsonii* from *Paspalum*). The example of *C. cereale*, where multiple well-supported phylogenetic lineages were ultimately categorized as populations after detailed sampling and extended analyses (Crouch *et al.* 2008, chapter 4), cautions against splitting these taxa without further study. The establishment of *C. nicholsonii* to encompass three phylogenetic species of *Colletotrichum* from *Paspalum* spp. gives rise to a monophyletic group. By not addressing the taxonomy of the two lineages of *Colletotrichum* associated with *S. officinarum* identified in this work, however; *C. falcatum* is preserved as a polyphyletic taxon, but its division without the examination of type material and more extensive study is unwise. For *C. falcatum* in particular, an important pathogen of sugarcane in developing nations, in-depth

population studies should first be performed to test the veracity of the multiple species scenario presented in this work, regardless of the temporary cladistical inconsistency.

For the six novel species proposed and described herein, along with *C. cereale*, *C. sublineolum*, *C. falcatum*, *C. caudatum* and *C. gramincola*, each taxon is characterized by their molecular identities at the *ITS1/5.8S/ITS2*, *Mat1-2*, *Apn1* and *Sod-2* loci. As previously noted, most of the FG *Colletotrichum* are morphologically cryptic, with no feature -- either individually or in combination -- capable of uniquely diagnosing these entities (this study; Crouch *et al.*. 2006 and 2008; chapter 1 and 4; Sutton 1980 and 1992). The distinctive appressorial size and shape of *C. graminicola* are an exception to this generalization, but even this character is not inviolable, as some *C. graminicola* individuals produce smaller, more irregularly shaped structures (Fig. 6.8.6); furthermore, the appressoria of *C. sublineolum* often mimic the appearance of *C. graminicola* appressoria (Fig. 6.8.5 and 6.8.6). To my knowledge, the only novel and diagnostic morphological feature of the FG group is the conidial appendage carried by isolates of *C. caudatum* (Sutton 1980).

Because the FG group is an assemblage of species with a relatively high level of host/species correspondence, host range provides a secondary character for purpose of identification of FG taxa. In some instances host range may prove unreliable for diagnostic purposes given the wide host range of some graminicolous *Colletotrichum* (e.g. *C. cereale* and *C. caudatum*), the possible existence of undescribed species, the possible existence of novel host association not described here or elsewhere in the literature (e.g. the association of *C. sublineolum* with *Eremochloa ophiuroides*; J.A. Crouch, unpublished data), and the potential for multiple species occurring on a single host species or genera (e.g. *S. officianarum* or *Paspalum* spp.). Used in combination, host range and appressorial

characteristics could form the basis of a satisfactory dichotomous key for identification of FG group species as described in this study.

6.4.1 ***Colletotrichum hanaui*** Crouch, J.F. White, B.B. Clarke, and Hillman, *sp. nov.* Figures 6.1, 6.5.2, 6.8.1, 6.11.9, 6.11.10 and 6.11.11.

Coloniae variabiles, albae vel griseo-brunneae, plerumque in massulas conidiales salmoneas dispositae. Setae septatae, fusco-brunneae rotundae, interdum ad basem lobatae, 75.0-25.0 um x 6.0-8.0 um, ad apicem acutae. Hyphae septatae, hyalinae, 1.0-6.5 um, interdum guttulatae. Conidia falcata vel fusiformia, ad apicem acuta vel obtusa, 21.5-25.5 um x 4.5-6 um; numero medio 23.7 um x 5.3 um. Appressoria hyphopodialia laevia, globosa vel prolata, ovoidea vel obovoidea ad apicem obtuse vel cylindracea, marginibus integris, 10.0 x 15.0 um x 5.5-7.5 um; numero medio 12.7 um x 5.5 um. Morphologia speciebus graminicolis *Colletotrichi* ceteris similis, sed per polymorphismos nucleotidicos ad genas ITS, *Sod2*, *Apn1* et *Mat1* atque consociatione peculiari cum graminibus generis *Digitariae* differt. Cognitio singulariter ex phylogenia per datorum ordinationem consequentiae nucleotidicae combinatae ITS, *Sod2*, *Apn1* et *Mat1* exorta.

Holotypus: Colonia sicca MAFF305404.

Colonies variable, white or grey-brown sometimes with salmon conidial masses. Setae septate, dark brown, rounded and sometimes lobed at base, 75.0-125.0 x 6.0-8.0 μm, apice acute. Hyphae septate, hyaline, 1.0-6.5 μm, sometimes guttulate. Conidia falcate

or fusiform, apices acute or obtuse, 21.5-25.5 x 4.5-6 μm; average 23.7 μm x 5.3 μm. Hyphopodial appressoria smooth, globose to prolate, ovoid or obovoid with obtuse or cylindrical apice, edges entire, 10.0 x 15.0 x 5.5-7.5 μm; average 12.7 x 5.5 μm. Morphologically similar to other grass-associated *Colletotrichum*, but differs based upon nucleotide polymorphisms at ITS, *Sod2*, *Apn1* and *Mat1* genes, and unique association with grasses of the genus *Digitaria*. Uniquely diagnosed through phylogeny based on combined ITS, *Sod2*, *Apn1* and *Mat1* nucleotide sequence dataset.

- **Type specimen:** MAFF305404, a strain originally isolated from *Digitaria ciliaris* from the Tochigi Prefecture in 1975; strain available from Genebank, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan.
- **Specimens examined:** JAPAN: MAFF305404 *Digitaria ciliaris*; MAFF511114, *D. ciliaris*. UNITED STATES: 1040DS, *Digitaria sp*.; NORTH CAROLINA: BPI 398389, *D. sanguinalis*, C.L. Lefebvre,1946-Aug-07, BPI; NORTH CAROLINA: BPI 398390, *D. sanguinalis*, C.L. Lefebvre, 1944-Sep-12, BPI; NORTH DAKOTA: BPI 398392A, *D. sanguinalis*, R. Sprague R., 1941-Sep25, BPI; NORTH DAKOTA: BPI 398392B, *D. sanguinalis*, R. Sprague R., 1941-Sep25, BPI.
- **Habitat**: Grasses of the genus *Digitaria*, including *D. ciliaris*.
- **Teleomorph**: Unknown.
- **Etymology.** Latinized from "hanau", referring to Dr. Robert Hanau, in recognition of his pioneering efforts in the field of fungal molecular biology and

the development of *C. graminicola* as a model system for the study of fungal molecular biology, pathogenicity, host recognition and genetics.

6.4.2 *Colletotrichum nicholsonii* Crouch, J.F. White, B.B. Clarke, and Hillman, *sp. nov.* Figures 6.1, 6.5.1 and 6.8.4.

Coloniae albae vel cremeo-albae, setis nullis. Hyphae septatae, hyalinae, 1.0-6.5 um, interdum guttulatae. Setae septatae, fusco-brunneae, rotundatae atque interdum ad basem flexae vel basirameae, 70.0-137.5 x 6.0-8.0 um, ad apicem acutae. Conidia falcata vel fusiformia, ad apicem peracuta, 17.5-23.5 x 3.5-5.0 um; numero medio 19.9 x 4.3 um. Appressoria hyphopodialia globosa vel perprolata, ovoidea vel obovoidea vel clavata, laevia vel lobata vel multi-lobata, ad apicem cylindrica vel obtusa, marginibus integris, 11.25-17.5 x 5.0-10.0 um; numero medio 12.7 x 9.4 um. Morphologia speciebus graminicolis generis *Colletotrichi* ceteris similis, sed per polymorphismos nucleotidicos ad genas ITS, *Sod2*, *Apn1* et *Mat1* atque consociatione peculiari cum graminibus generis *Paspali* differt. Cognitio singulariter ex phylogenia per datorum ordinationem consequentiae nucleotidicae combinatae ITS, *Sod2*, *Apn1* et *Mat1* exorta.

Holotypus: Colonia sicca MAFF511115.

Colonies white or creamy white, setae absent. Hyphae septate, hyaline,1.0-6.5 μm sometimes guttulate. Setae septate, dark brown, rounded and sometimes bent or branching at base, 70.0-137.5 x 6.0-8.0 μm, apice acute. Conidia falcate or fusiform,

apices sharply acute, 17.5-23.5 x 3.5-5.0 μm; average 19.9 x 4.3 μm. Hyphopodial appressoria globose to perprolate, ovoid or obovoid or clavate, smooth or lobate or multi-lobate, apice cylindrical or obtuse, edges entire, 11.25-17.5 x 5.0-10.0 μm; average 12.7 x 9.4 μm. Morphologically similar to other grass-associated *Colletotrichum*, but differs based upon nucleotide polymorphisms at ITS, *Sod2*, *Apn1* and *Mat1* genes, and unique association with grasses of the genus *Paspalum*. Uniquely diagnosed through phylogeny based on combined ITS, *Sod2*, *Apn1* and *Mat1* nucleotide sequence dataset.

- **Type specimen:** JAPAN: MAFF511115, a strain originally isolated from *P. dilatatum* from an unknown location; strain available from Genebank, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan.
- **Specimens examined**: JAPAN: MAFF305403, *Pasplaum notatum*, 1977; MAFF305428, *P. notatum*, 1977; MAFF511115, *P. dilatatum*; MAFF510916, *P. dilatatum*, 1975,; MAFF305391, *P. dilatatum* 1974,; MAFF511000, *P. notatum*, 1975,; UNITED STATES: 1047, 1985, *Colletotrichum* Germplasm Database; and NEW ZEALAND: 5171, *Paspalum* spp., 1965, ICMP.
- **Habitat**: Grasses of the genus *Paspalum*, including *P. dilatatum* and *P. notatum*.
- **Teleomorph**: Unknown.
- **Etymology.** Latinized from "'Nicholson", referring to Dr. Ralph Nicholson, in recognition of his significant contributions to our understanding of how fungi survive adversity through studies using *C. graminicola* as a model system.

6.4.3 ***Colletotrichum jacksonii*** Crouch, J.F. White, B.B. Clarke, and Hillman, *sp. nov.* Figures 6.1, 6.5.2, 6.8.2, 6.11.1, 6.11.2, 6.11.3, 6.11.4, 6.11.5.

Coloniae variabiles, cremeae et brunellae vel mediocriter brunneae et brunellae, interdum in massas conidiales salmoneas dispositae. Setae septatae, fusco-brunneae, ad basem rotundatae, 87.5-125.0 x 6.0-8.0 um, ad apicem acutae. Hyhae septatae, hyalinae, 1.0-6.5 um, saepe guttulatae. Conidia falcata vel fusiformia, ad apicem acuta vel peracuta, 18.5-23.5 x 3.5-4.0 um; numero medio 19.9 x 3.7 um. Appressoria hyphopodialia globosa vel perprolata, ovoidea vel obovoidea vel clavata, laevia vel lobata vel multilobata, ad apicem cylindrica vel obtusa, marginibus integris, 8.75-15.0 x 7.5-12.5 um; numero medio 12.3 x 11.1 um. Morphologia speciebus graminicolis generis *Colletotrichi* ceteris similis, sed per polymorphismos nucleotidicos ad genas ITS, *Sod2*, *Apn1* et *Mat1* atque consociatione peculiari cum graminibus generis *Echinochloae* differt. Cognitio singulariter ex phylogenia per datorum ordinationem consequentiae nucleotidicae combinatae ITS, *Sod2*, *Apn1* et *Mat1* exorta.

Holotypus: Colonia sicca MAFF305460.

Colonies variable, creamy white and tan or medium brown with tan, sometimes with salmon conidial masses. Setae septate, dark brown, rounded at base, 87.5-125.0 x 6.0-8.0 μm, apice acute. Hyphae septate, hyaline,1.0-6.5 μm, often guttulate. Conidia falcate or fusiform, apices acute or sharply acute, 18.5-23.5 x 3.5-4.0 μm; average 19.9 x 3.7 μm. Hyphopodial appressoria globose to perprolate, ovoid or obovoid or clavate, smooth or

lobate or multi-lobate, apice cylindrical or obtuse, edges entire, 8.75-15.0 x 7.5-12.5 μm; average 12.3 x 11.1 μm. Morphologically similar to other grass-associated *Colletotrichum*, but differs based upon nucleotide polymorphisms at ITS, *Sod2*, *Apn1* and *Mat1* genes, and unique association with grasses of the genus *Echinochloa*. Uniquely diagnosed through phylogeny based on combined ITS, *Sod2*, *Apn1* and *Mat1* nucleotide sequence dataset.

- **Type specimen:** JAPAN: MAFF305460, a strain originally isolated from *echinochloa esculenta* in the Tochigi Prefecture in 1980; strain available from Genebank, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan.
- **Specimens examined**: JAPAN: MAFF305460, *Echinochloa esculenta*, 1980, MAFF305439, *E. esculenta*, 1977; MAFF511152, *E. esculenta*, 1977; MAFF511328, *E. esculenta*, 1980;; and MAFF511344, *E. esculenta*, 1985; UNITED STATES: MARYLAND: BPI 398394, *E. crus-galli*, C.L. Lefebvre, 1941-Aug-27, BPI; GEORGIA: BPI 398395, *E. crus-galli*, 1943-Sep-25, BPI; WISCONSIN: BPI 398396, *E. crus-galli*, A.G. Johnson, BPI; 1912-Aug-12; MARYLAND: BPI 398406, *E. pungens*, F.J. Hermann, 1945-Aug-26, BPI.
- **Habitat**: Grasses of the genus *Echinochloa*, including *E. esculenta*.
- **Teleomorph**: Unknown.
- **Etymology.** Latinized from Jackson, referring to Dr. Wes Jackson, founder of The Land Institute, in recognition of his profound contributions to agriculture, science, society and education.

6.4.4 ***Colletotrichum miscanthi*** Crouch, J.F. White, B.B. Clarke, and Hillman, *sp. nov.* Figure 6.1 and 6.5.2.

Coloniae crassae, mycelio pallide griseo annulis mediocriter griseis atque marginibus irregulariter pannosis ornato. Hyphae septatae, hyalinae, 1.0 -6.5 um, saepe guttulatae. Setae septatae, fusco-brunneae, ad basem rotundatae, 75.0-125.0 x 6.0 -8.0 um, ad apicem acutae. Conidia falcata vel fusiformia, ad apicem acuta, 19.5-25.5 x 4 -4.5 um; numero medio 23.2 x 4.1 um. Appressoria hyphopodialia nulla. Morphologia speciebus graminicolis generis *Colletotrichi* ceteris similis, sed per polymorphismos nucleotidicos ad genas ITS, *Sod2*, *Apn1* et *Mat1* atque consociatione peculiari cum graminibus generis *Miscanthi* differt. Cognitio singulariter ex phylogenia per datorum ordinationem consequentiae nucleotidicae combinatae ITS, *Sod2*, *Apn1* et *Mat1* exorta.

Holotypus: Colonia sicca MAFF510857.

Colony thick pale grey mycelium with medium grey rings and irregular felty edge. Hyphae septate, hyaline, 1.0-6.5 μm, often guttulate. Setae septate, dark brown, rounded at base, 75.0-125.0 μm x 6.0-8.0 μm, apice acute. Conidia falcate or fusiform, apices acute, 19.5-25.5 μm x 4-4.5 μm; average of 23.2 μm x 4.1 μm. No hyphopodial appressoria produced. Morphologically similar to other grass-associated *Colletotrichum*, but differs based upon nucleotide polymorphisms at ITS, *Sod2*, *Apn1* and *Mat1* genes, and unique association with grasses of the genus *Miscanthus*. Uniquely diagnosed through phylogeny based on combined ITS, *Sod2*, *Apn1* and *Mat1* nucleotide sequence dataset.

- **Type specimen:** JAPAN: MAFF510857, a strain originally isolated from the Tochigi Prefecture in 1972; strain available from Genebank, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan.
- **Habitat**: Grasses of the genus *Miscanthus*, including *M. sinensis*.
- **Teleomorph**: Unknown.
- **Etymology.** "Of *Miscanthus*".

6.4.5 ***Colletotrichum axonopodi*** Crouch, J.F. White, B.B. Clarke, and Hillman, *sp. nov.* Figures 6.1, 6.8.3, 6.11.7 and 6.11.8.

Coloniae myceliis albis super mycelium brunneum et brunello-aurantiacum pallide tinctum positis. Setae septatae, fusco-brunneae, ad basem rotundatae, 65.0-125.0 x 6.0-8.0 um, ad apicem acutae. Hyphae septatae, hyalinae, 1.0-6.5 um, saepe guttulatae. Conidia falcata vel fusiformia, ad apicem acuta, 20.0-25.0 x 3.5-5.0 um; numero medio 24.2 um x 4.2 um. Appressoria hyphopodialia globosa vel perprolata, ovoidea vel obovoidea vel clavata, laevia vel lobata vel multilobata, ad apicem cylindrica vel obtusa, marginibus integris, 7.5-11.25 x 5.0-12.5 um, numero medio 10.1 x 7.5 um. Morphologia speciebus graminicolis generis *Colletotrichi* ceteris similis, sed per polymorphismos nucleotidicos ad genas ITS, *Sod2*, *Apn1* et *Mat1* atque consociatione peculiari cum graminibus generis *Axonopodis* differt. Cognitio singulariter ex phylogenia per datorum ordinationem consequentiae nucleotidicae combinatae ITS, *Sod2*, *Apn1* et *Mat1* exorta.

Holotypus: Colonia sicca IMI279189.

Colonies with white mycelia overlaid on brown mycelium with pale tan-orange cast. Setae septate, dark brown, rounded at base, 65.0-125.0 μm x 6.0-8.0 μm, apice acute. Hyphae septate, hyaline1.0-6.5 μm, often guttulate. Conidia falcate or fusiform, apices acute, 20.0-25.0 μm x 3.5-5.0 μm; average of 24.2 μm x 4.2 μm. Hyphopodial appressoria globose to perprolate, ovoid or obovoid or clavate, smooth or lobate or multi-lobate, apice cylindrical or obtuse, edges entire, 7.5-11.25 μm x 5.0-12.5 μm average 10.1 μm x 7.5μm. Morphologically similar to other grass-associated *Colletotrichum*, but differs based upon nucleotide polymorphisms at ITS, *Sod2*, *Apn1* and *Mat1* genes, and unique association with grasses of the genus *Axoponuss*. Uniquely diagnosed through phylogeny based on combined ITS, *Sod2*, *Apn1* and *Mat1* nucleotide sequence dataset.

- **Type specimen:** AUSTRALIA: IMI279189, a strain originally isolated from *A. affinis* in Queensland in 1983; strain available from CABI Europe - UK, Bakeham Lane, Egham, Surrey TW20 9TY, UK.
- **Specimens examined:** UNITED STATEES: LOUISIANA: BPI 398338; *A. affinis*, C.L. Lefebvre, 1940-Sep-13, BPI; GEORGIA: BPI 398340, *A. compressus*, H.W. Johnson, 1930-May-20, BPI; HONDURAS: BPI 398339, *A. compressus*, P.C. Standley, 1946-Dec-10, BPI.
- **Habitat**: Grasses of the genus *Axonopus*, including *A. affinis*.
- **Teleomorph**: Unknown.
- **Etymology.** Latinized from axonopus: “from axonopus”.

6.4.6 ***Colletotrichum eleusines*** Pavgi & U.P. Singh [as 'eleusinis'], Mycopath. Mycol. appl. 27: 85 (1965); Figures 6.1, 6.5.2, 6.8.3, 6.11.6.

Emended description: Colonies with white mycelia overlaid on brown mycelium with intense creamy, pale orange cast; setae absent; conidia sparse. Setae septate, dark brown, rounded at base, 65.0-125.0 μm x 6.0-8.0 μm, apice acute. Hyphae septate, hyaline, 1.0-6.5 μm, often guttulate. Conidia falcate or fusiform, apices acute, 24.5-28.5 μm x 3.5-4.0 μm; average of 26.4 μm x 3.7 μm. Hyphopodial appressoria globose to perprolate, ovoid or obovoid or clavate, smooth, rarely lobate, apice cylindrical or obtuse, edges entire, 10.0-13.75 μm x 7.5-10.0 μm; average 12.6 μm x 8.7 μm. Morphologically similar to other grass-associated *Colletotrichum*, but differs based upon nucleotide polymorphisms at ITS, *Sod2*, *Apn1* and *Mat1* genes, and unique association with grasses of the genus *Eleusine*. Uniquely diagnosed through phylogeny based on combined ITS, *Sod2*, *Apn1* and *Mat1* nucleotide sequence dataset.

- **Specimens examined:** JAPAN: MAFF511155; UNITED STATES: FLORIDA: II488, *E. indica*, G.W. Weber, 1936-July-07, NY.
- **Epitype establishment**: To facilitate species interpretation, MAFF511155 from *E. indica* is designated as an epitype; a culture of this strain is maintained at at Genebank, National Institute of Agrobiological Sciences, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan. This strain was originally isolated from the Kumamoto Prefecture, Japan in 1977.
- **Teleomorph**: Unknown

6.5 References

Cannon, P. F., Bridge, P. D., and Monte, E. 2000. Linking the past, present, and future of *Colletotrichum* systematics. Pp. 1-20 in: Colletotrichum: Host specificity, pathology, and host-pathogen interaction. D. Prusky, S. Freeman and M. B. Dickman (eds.). APS Press, St. Paul, MN.

Crouch, J.A., Clarke, B.B. and Hillman, B.I. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology **96**: 46-60.

Crouch, J.A., Tredway, L.P., Clarke, B.B. and Hillman, B.I. 2008(a). Ecological specialization drives the evolution of the pathogenic fungus *Colletotrichum cereale* and allied taxa across diverse grass communities. In review with Mol. Ecol.

Crouch, J.A., Glasheen, B.M., Giunta, M.A., Clarke, B.B. and Hillman, B.I. 2008(b). The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. Fungal Genetics Biol. 45:190–206.

Crouch, J.A., Glasheen, B.M., Uddin, W., Clarke, B.B., and Hillman, B.I. (2008c) Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles. In press, Crop Science.

Dettman JR, Jacobson DJ, Taylor JW (2003) A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* **57**, 2703-2720.

Du, M., Schardl, C. L., Nuckles, E. M., and Vaillancourt, L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Colletotrichum* species complexes.

Farr, D.F., Aime, M.C., Rossman, A.Y. and Palm, M.E. 2006. Species of *Colletotrichum* in agavaceae. Mycol. Res. 110: 1395-1408.

Huelsenbeck, J. P., and Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Posada, D., and Crandall, K. A. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Sutton, B. C. 1965. Studies on the taxonomy of *Colletotrichum* Cda with especial reference to *C. graminicola* (Ces.) Wilson, University of London, London.

Sutton, B. C. 1966. Development of fruitifications in *Colletotrichum graminicola* (Ces.) Wils. and related species. Can. J. Bot. 44:887-897.

Sutton, B. C. 1968. The appressoria of *Colletotrichum graminicola* and *C. falcatum*. Can. J. Bot. 46:873-876.

Sutton, B. C. 1980. The coelomycetes: fungi imperfecti with pycnidia, acervuli, and stromata. Kew, U.K.: Commonwealth Mycological Institute.

Sutton, B. C. 1992. The genus *Glomerella* and its anamorph *Colletotrichum*. Pp. 1-26 *In Colletotrichum*: Biology, pathology and control. J. A. Bailey and M. J. Jeger (eds.). CAB International, Wallingford, U.K.

Swofford, D. L. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S., and Fisher, M. C. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31:21-32.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. Nucleic Acids Res. 22:4673-4680.

Vaillancourt, L., Du, M., Wang, J., Rollins, J., and Hanau, R. 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia 92:430-435.

Table 6.1

The 55 strains of *Colletotrichum* isolated from grass hosts used in this study

Isolate name	Species	Host plant species	Country of origin	Other available origination data	Year	Source	Pop.
279189-AA	*C. axonopodus*	*Axonoponus affinis*	Australia	Queensland	1983	IMI279189	—
176617-BB	*C. caudatum*	*Bothriochloa bladhii*	Australia	—	1973	IMI 176617	—
176619-IC	*C. caudatum*	*Imperata cylindrica*	Australia	Caboolture	1973	IMI 176619	—
305700-IC	*C. caudatum*	*Imperata cylindrica*	Japan			MAFF 305700	—
238575-ZT	*C. caudatum*	*Zoysia tennufolia*	Japan			MAFF 238575	—
24049-AS	*C. cereale*	*Avena sativa*	Germany	—	1949	CBS 240.49	A1
CA-ANCG17-14	*C. cereale*	*Poa annua*	USA	Pasedena, CA	2004	—	A10
305377AE	*C. cereale*	*Arrhenatherum elatius*	Japan	Chiba Prefecture	1967	MAFF 510634	A2
KS-TA-4-F4	*C. cereale*	*Triticum aestivum*	USA	Marshall County, KS	2005	—	A3
305429-PF	*C. cereale*	*Polypogon fugax*	Japan	Saga Prefecture	1977	MAFF 305429	A4
305076-AS	*C. cereale*	*Avena sativa*	Japan	Saga Prefecture	1966	MAFF 305076	A5
IL-CI-7.3D	*C. cereale*	*Calamagrostis inexpansa*	USA	Markham, IL	2005	—	A6
NJ-8626	*C. cereale*	*Poa annua*	USA	Middletown, NJ	2004	—	A7
68188-lg	*C. cereale*	"lawn grass"	Netherlands	—	1988	CBS 681.88	A8
NJ-HF2B	*C. cereale*	*Poa annua*	USA	New Brunswick, NJ	2003	—	A9
NJ-CA1C1	*C. cereale*	*Calamagrostis acutifolia*	USA	Barrington, NJ	2005	—	B
NJ-DG-2A2-5	*C. cereale*	*Dactylis glomerata*	USA	Sussex County, NJ	2005	—	B
511155-EI	*C. eleusines*	*Eleusine indica*	Japan	Kumamoto Prefecture	1977	MAFF 511155	—
78362-SO	*C. falcatum*	*Saccarum officinarum*	Nigeria		1960	IMI 78362	—
16970-SO	*C. falcatum*	*Saccharum officinarum*	Brazil		1970	CBS 169.70	—
24362-SO	*C. falcatum*	*Saccharum officinarum*	Brazil		1962	CBS 243.62	—
305077-SO	*C. falcatum*	*Saccharum officinarum*	Japan	Chiba Prefecture	1966	MAFF 305077	—
306170-SO	*C. falcatum*	*Saccharum officinarum*	Japan			MAFF 306170	—
306299-SO	*C. falcatum*	*Saccharum officinarum*	Japan			MAFF 306299	—
347765-SO	*C. falcatum*	*Saccharum officinarum*	Nigeria		—		—
M1001	*C. graminicola*	*Zea mays*	USA	Missouri	1978	—	—
IN-900190	*C. graminicola*	*Zea mays*	USA	Indiana	1990	—	—
IN-12475	*C. graminicola*	*Zea mays*	USA	Indiana	1975	—	—
MO-178	*C. graminicola*	*Zea mays*	USA	Missouri	1978	—	—
KY-197	*C. graminicola*	*Zea mays*	USA	McClean County, IN	1997	—	—
NY-15182	*C. graminicola*	*Zea mays*	USA	Tioga Co., NY	1982	—	—
27554-ZM	*C. graminicola*	*Zea mays*	Netherlands	—	1954	CBS 275.54	—
311343-ZM	*C. graminicola*	*Zea mays*	Japan	—	1985	MAFF 311343	—
305404-DC	*C. hanaui*	*Digitaria ciliaris*	Japan	Tochigi Prefecture	1975	MAFF 305404	—
511014-DC	*C. hanaui*	*Digitaria ciliaris*	Japan	Tochigi Prefecture	1975	MAFF 511014	—
1040-DS	*C. hanaui*	*Digitaria* sp.	USA	Monticello, AR	1992	CDG 1040	—
1391-DS	*C. hanaui*	*Digitaria* sp.	USA			CDG 1391	—
305439-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Miyazaki Prefecture	1977	MAFF 305439	—
305460-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Tochigi Prefecture	1980	MAFF 305460	—
511152-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Kochi Prefecture	1977	MAFF 511152	—
511328-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Tochigi Prefecture	1980	MAFF 511328	—
511344-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Tochigi Prefecture	1985	MAFF 511344	—

Table 6.1, continued

Isolate name	Species	Host plant species	Country of origin	Other available origination data	Year	Source	Pop.
510857-MS	*C. miscanthi*	*Miscanthus sinensis*	Japan	Tochigi Prefecture	1972	MAFF 510857	—
1047-PD	*C. nicholsonii*	*Paspalum dilatatum*	USA	Baldwin Springs, AR	1985	CDG 1047	—
305391-PD	*C. nicholsonii*	*Paspalum dilatatum*	Japan	Chiba Prefecture	1974	MAFF 305391	—
510916-PD	*C. nicholsonii*	*Paspalum dilatatum*	Japan	Chiba Prefecture	1975	MAFF 510916	—
511115-PD	*C. nicholsonii*	*Paspalum dilatatum*	Japan			MAFF 511115	—
305403-PN	*C. nicholsonii*	*Paspalum notatum*	Japan	Yamaguchi Prefecture	1977	MAFF 305403	—
305428-PD	*C. nicholsonii*	*Paspalum notatum*	Japan	Kumamato Prefecture	1977	MAFF 305428	—
511000-PN	*C. nicholsonii*	*Paspalum notatum*	Japan		1975	MAFF 511000	—
5171-PS	*C. nicholsonii*	*Paspalum* sp.	New Zealand	Auckland	1965	ICMP 5171	—
S3001	*C. sublineolum*	*Sorghum bicolor*	Burkina Fasso		—	—	—
305360-SB	*C. sublineolum*	*Sorghum bicolor*	Japan		1957	MAFF 305360	—
510021-SB	*C. sublineolum*	*Sorghum bicolor*	Japan		1957	MAFF 510021	—
TX-BI2K	*C. sublineolum*	*Sorghum halapense*	USA	Brazoria County, TX	2005	—	—

Table 6.2

Colletotrichum isolated from non-graminicolous hosts used in this study

Isolate name	Species	Host plant	Location	Location	Year	Source
33571-CA	*C. capsici*	*Capsicum annuum*	Indonesia	Java, Bogor	1975	CBS 335.75
603-DS	*C. capsici*	*Datura stramonium*	USA	California	1986	CDG 603
101631-HS	*C. capsici*	*Hemerocallis* sp.	New Zealand	Whenuapai, Auckland	1999	CBS 101631
117546-AP	*C. circinans*	*Allium porrum*	Netherlands	--	2004	CBS 117546
366-AV	*C. circinans*	*Allium vineale*	USA	Greathouse Springs, AR	1984	CDG 366
23699	*C. dematium*		Japan			MAFF 23699
237705	*C. dematium*		Japan			MAFF 237705
305700	*C. dematium*		Japan			MAFF 305700
410758	*C. dematium*		Japan			MAFF 410758
410759	*C. dematium*		Japan			MAFF 410759
840067	*C. dematium*		Japan			MAFF 840067
840068	*C. dematium*		Japan			MAFF 840868
840069	*C. dematium*		Japan			MAFF 840098
840866	*C. dematium*		Japan			MAFF 840866
11-AA	*C. dematium*	*Aeschynomene americana*	USA	Fort Pierce, FL	1982	CDG 11
74-AS	*C. dematium*	*Amaranthus* sp.	USA	Sumter Co., AR	1983	CDG 74
850-CS	*C. dematium*	*Crotalaria spectabilis*	USA	Gainesville, FL	1988	CDG 850
851-CS	*C. dematium*	*Crotalaria spectabilis*	USA	Gainesville, FL	1988	CDG 851
1339-FS	*C. dematium*	*Freycinetia* sp.	New Zealand	--	1992	CDG 1339
1172-LS	*C. dematium*	*Lilium* sp.	USA	Muncy, PA	1991	CDG 1172
18630-LS	*C. dematium*	*Lilium* sp.	Netherlands	--	1930	CBS 18630
119444-LM	*C. dematium*	*Liriope muscari*	Mexico	--	2000	CBS 119444
84-OS	*C. dematium*	*Oenothera* sp.	USA	Washington Co., AR	1983	CDG 84
202-PP	*C. dematium*	*Polygonum pennsylvanicum*	USA	Fayetteville, AR	1984	CDG 202
1072-PL	*C. dematium*	*Pueraria lobata*	USA	Lake Weddington, AR	1990	CDG 1072
1075-PL	*C. dematium*	*Pueraria lobata*	USA	Washington Co., AR	1990	CDG 1075
1163-ST	*C. dematium*	*Solanum tuberosum*	USA	Rhode Island	1971	CDG 1163
680-TV	*C. dematium*	*Trillium viridens*	USA	--	1986	CDG 680
681-TV	*C. dematium*	*Trillium viridens*	USA	--	1986	CDG 681
682-TV	*C. dematium*	*Trillium viridens*	USA	--	1986	CDG 682
64-XS	*C. dematium*	*Xanthium strumarium*	USA	Washington Co., AR	1983	CDG 64
66-XS	*C. dematium*	*Xanthium strumarium*	USA	Washington Co., AR	1983	CDG 66
11221-LU	*C. lini*	*Linum usitatissimum*	UK	--	1921	CBS 112.21
990-LS	*C. lini*	*Linum usitatissimum*	UK	--	1963	CDG 990
34168-BS	*C. trichellum*	*Bambusa* sp.	USA	Alabama	--	ATCC 34168

Table 6.2, continued

Isolate name	Species	Host plant	Location	Location	Year	Source
118198-HH	*C. trichellum*	*Hedera helix*	Guatemala	--	2002	CBS 118198
21764-HH	*C. trichellum*	*Hedera helix*	UK	Berkshire, Binfield	1964	CBS 217.64
415-HH	*C. trichellum*	*Hedera helix*	USA	Fayetteville, AR	1986	CDG 415
416-HH	*C. trichellum*	*Hedera helix*	USA	Fayetteville, AR	1985	CDG 416
71-HH	*C. trichellum*	*Hedera helix*	USA	Washington Co., AR	1983	CDG 71
305982-PE	*C. trichellum*	*Passiflora edulis*	Japan			MAFF 305982
1349-MS	*C. trifolli*	*Medicago sativa*	New Zealand	--	1992	CDG 1349
1456-GM	*C. truncatum*	*Glycine max*	USA	Fayetteville, AR	1993	CDG 1456
1399-LC	*C. truncatum*	*Lens culinaris*	Canada	--	1992	CDG 1399
1400-LC	*C. truncatum*	*Lens culinaris*	Canada	Outlook, Saskatchewen	1992	CDG 1400
1402-LC	*C. truncatum*	*Lens culinaris*	Canada	Portage La Prairie, Manitoba	1992	CDG 1402

Table 6.3

Hyphopodial appresssoria characteristics of the graminicolous *Colletotrichum*.

Species	Shape	Height (µm)		Width (µm)	
		Range	Average	Range	Average
C. axonopodi	Globose to perprolate, ovoid or obovoid or clavate, smooth or lobate or multi-lobate, apice cylindrical or obtuse, edges entire	7.5–11.5	10.1	5–12.5	7.5
C. caudatum		12.5–32.5	18.4	5–12.5	11.2
C. cereale *	Rounded or smooth or irregular or lobate or multi-lobate, apice XXX, edges XXX	7.5–13.5	12.4	6.0–10.5	9.6
C. eleusines	Globose to perprolate, ovoid or obovoid or clavate, smooth, rarely lobate, apice cylindrical or obtuse, edges entire	10–13.5	12.6	7.5–10.0	8.7
C. falcatum		12.5–15.0	13.1	8.5–12.5	10.9
C. graminicola		10.0–22.5	18.8	7.5–20.0	14.5
C. hanaui	Smooth, globose to prolate, ovoid or obovoid with obtuse or cylindrical apice, edges entire	10.0–15.0	12.7	4.5–11.2	6.9
C. jacksonii	Globose to perprolate, ovoid or obovoid or clavate, smooth or lobate or multi-lobate, apice cylindrical or obtuse, edges entire	8.75–11.5	12.3	9.0–12.5	11.1
C. miscanthi	ND	ND	ND	ND	ND
C. nicholsonii	Globose to perprolate, ovoid or obovoid or clavate, smooth or lobate or multi-lobate, apice cylindrical or obtuse, edges entire	11.25–17.5	12.7	5.0–10.0	9.4
C. sublineolum		10.0–21.5	16.5	10–17.0	14.8

- Note: the compound appressoria produced by *C. cereale* NJ-HF2B was not included in these measurements.
- ND = no data

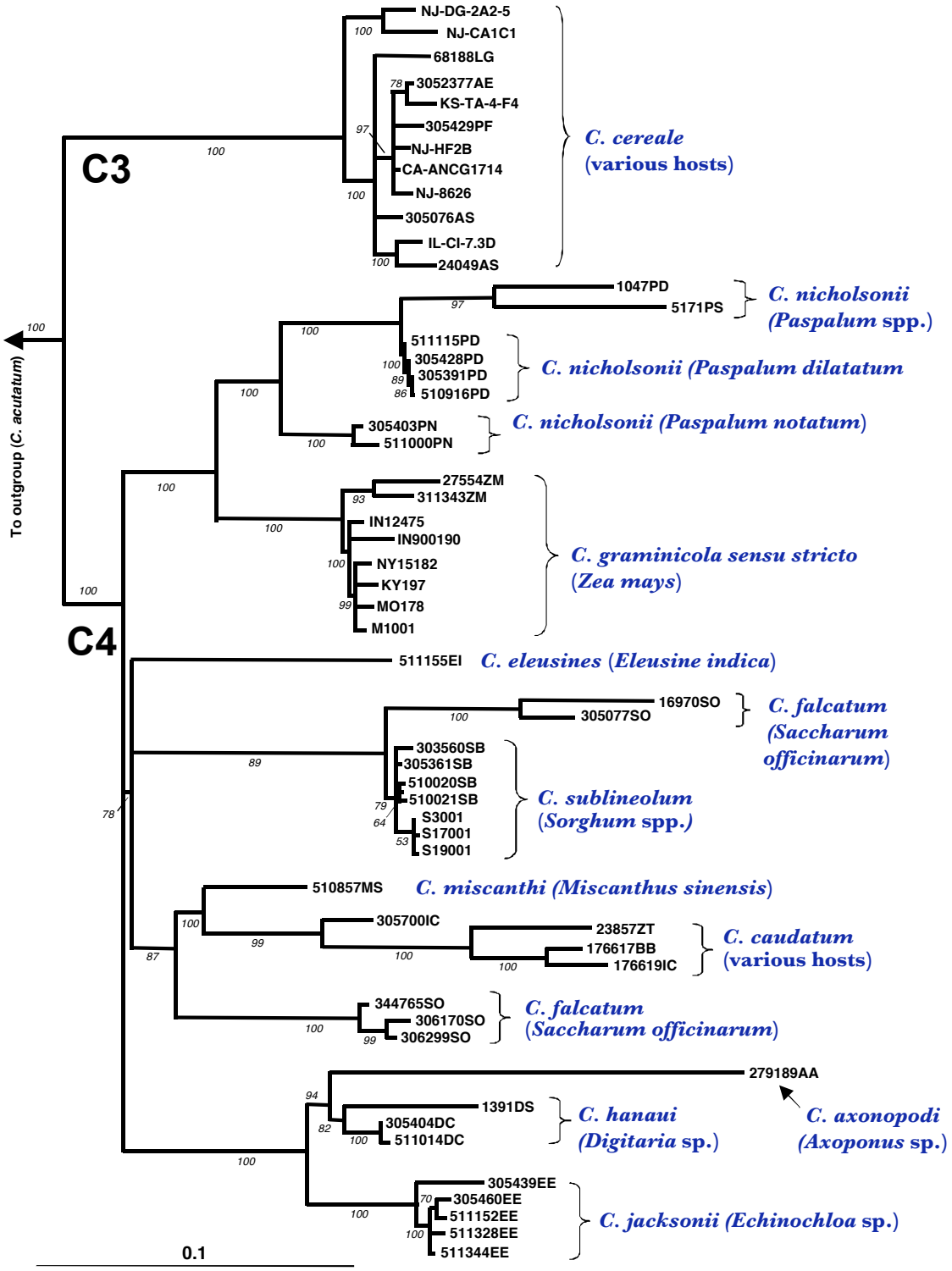


Figure 6.1
Phylogenetic tree constructed from the combined dataset.

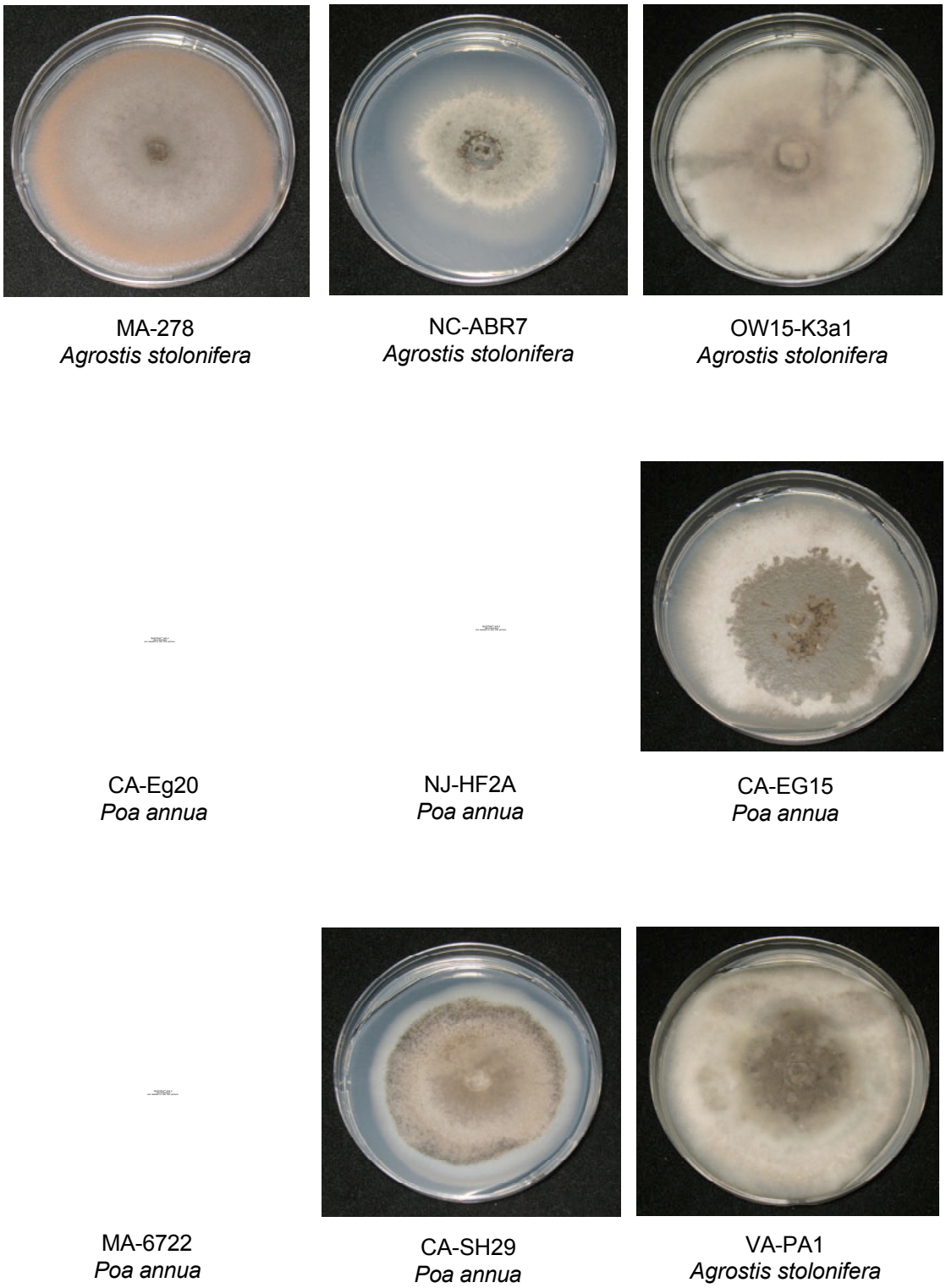
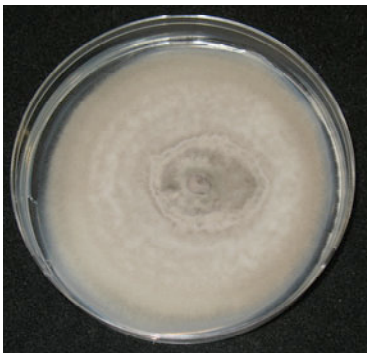


Figure 6.2.1
Isolates of *Colletotrichum cereale* growing on potato dextrose agar under continuous light, 5-8 d. Labels: First line of text is isolate name; second line indicates plant host. All pictured cultures were isolated from golf course turfgrass.

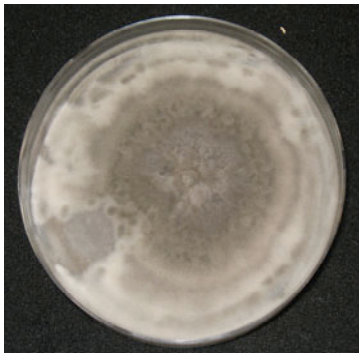


TN-GBGC5
Agrostis stolonifera

IL-PV2
Poa annua

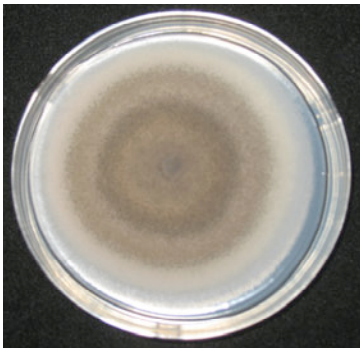
OW15-R1-3
Agrostis stolonifera

NJ-6553
Poa annua



ONT-99325
Agrostis stolonifera

ONT-00130
Agrostis stolonifera



OW15 F52
Agrostis stolonifera

OW15E302
Agrostos stolonifera

CT297
Agrostis stolonifera

Figure 6.2.2
Isolates of *Colletotrichum cereale* growing on potato dextrose agar under continuous light, 5-8 d. Labels: First line of text is isolate name; second line indicates plant host. All pictured cultures were isolated from golf course turfgrass.

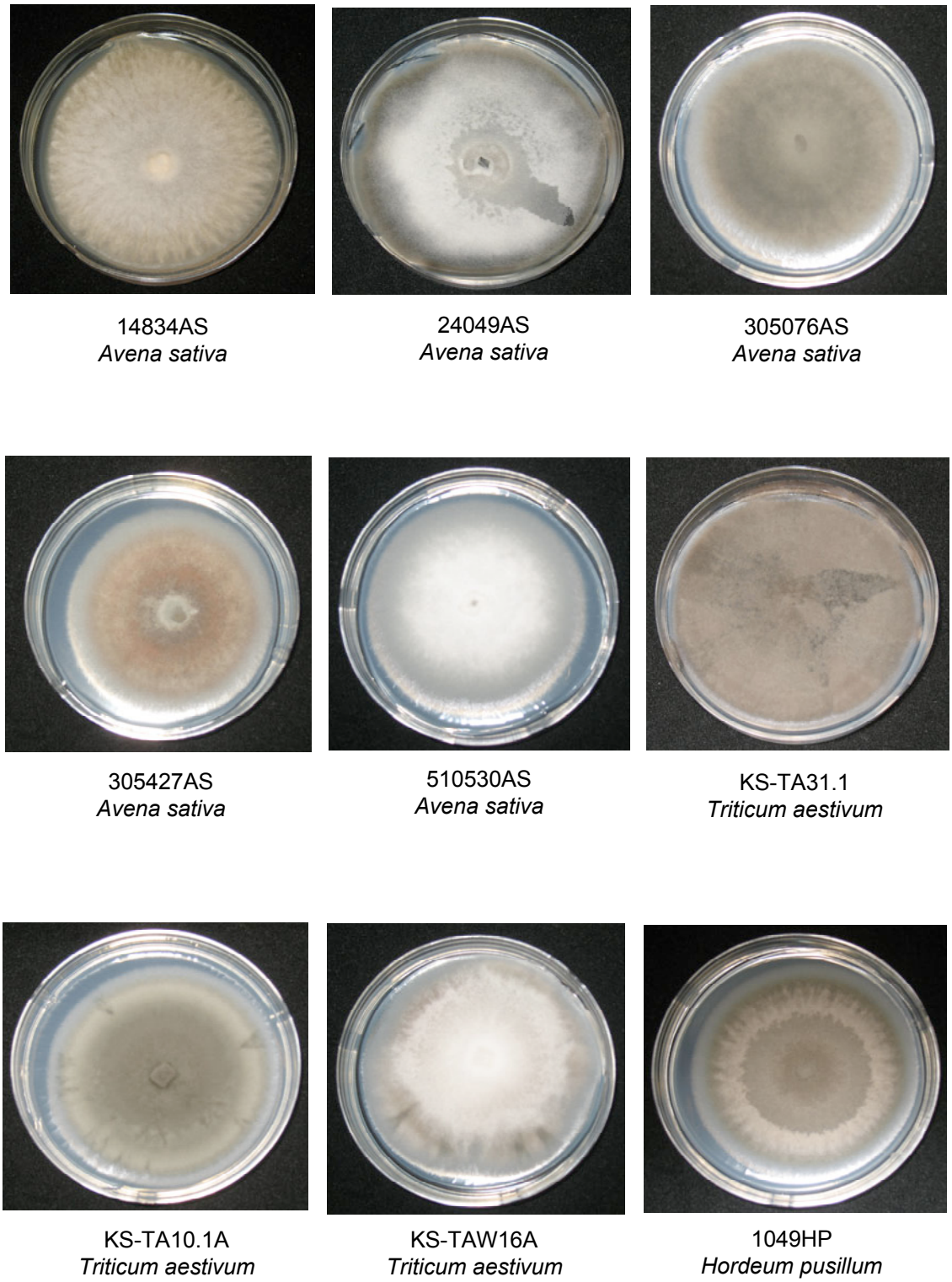


Figure 6.3.1
Isolates of *Colletotrichum cereale* growing on potato dextrose agar under continuous light, 5-8 d. All pictured cultures were isolated fromcool-season cereal crops: *Avena sativa* (oats), *Triticum aestivum* (wheat) and *Hordeum pusillum* (barley).

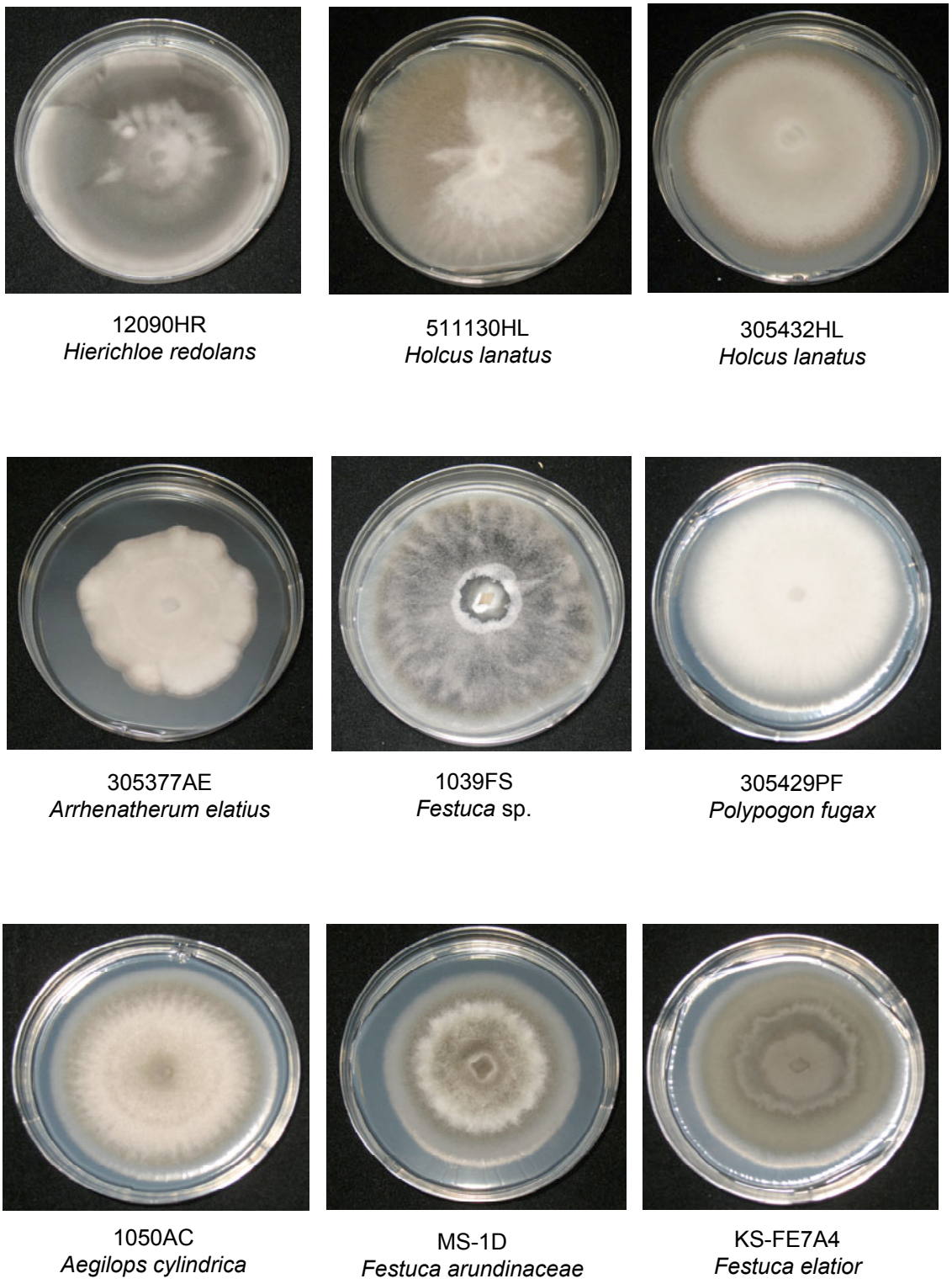


Figure 6.4.1
Isolates of *Colletotrichum cereale* isolated from various cool season grasses growing on potato dextrose agar under continuous light, 5-8 d.

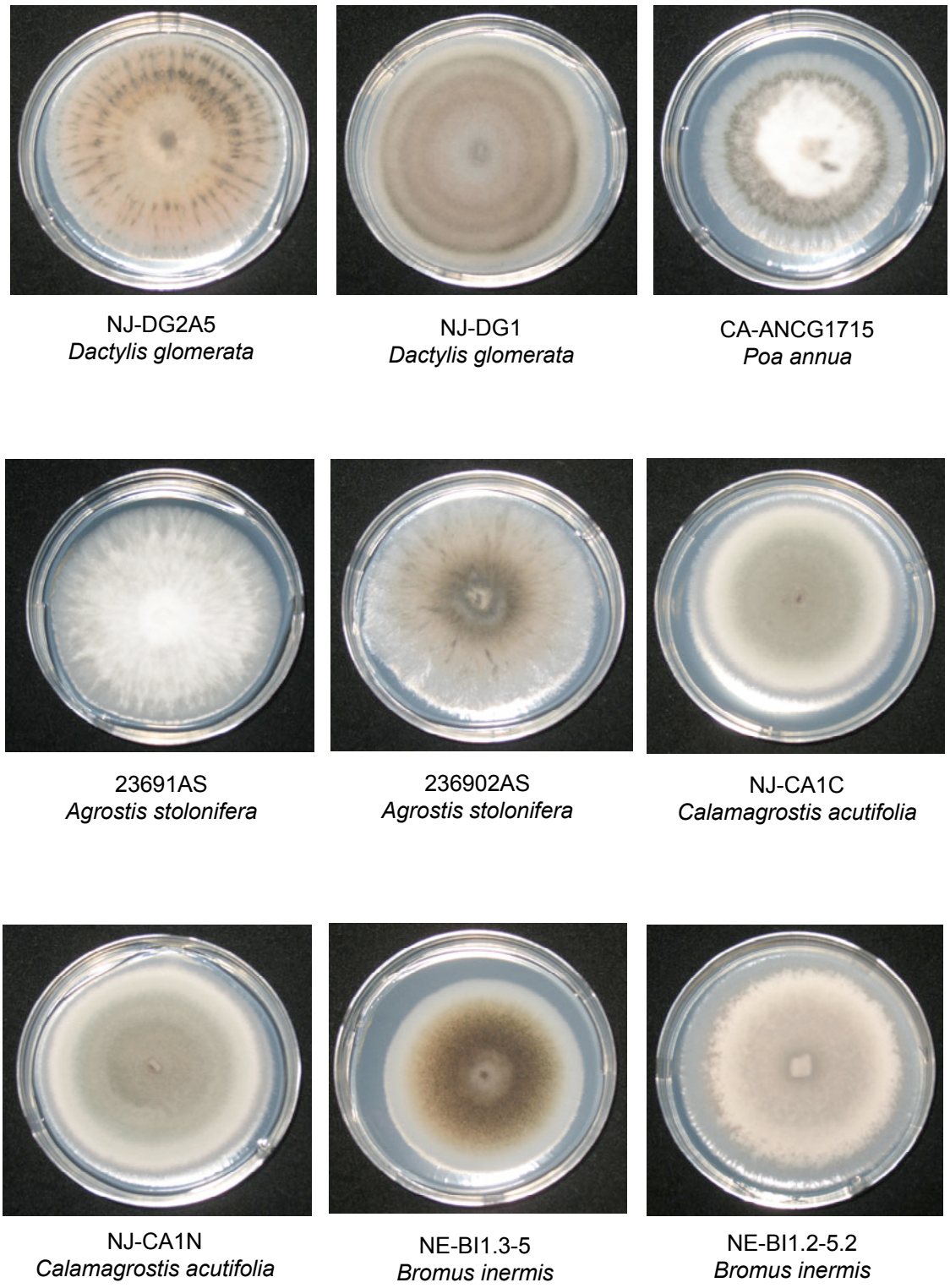


Figure 6.4.2
Isolates of *Colletotrichum cereale* isolated from various cool-season grasses growing on potato dextrose agar under continuous light, 5-8 d.

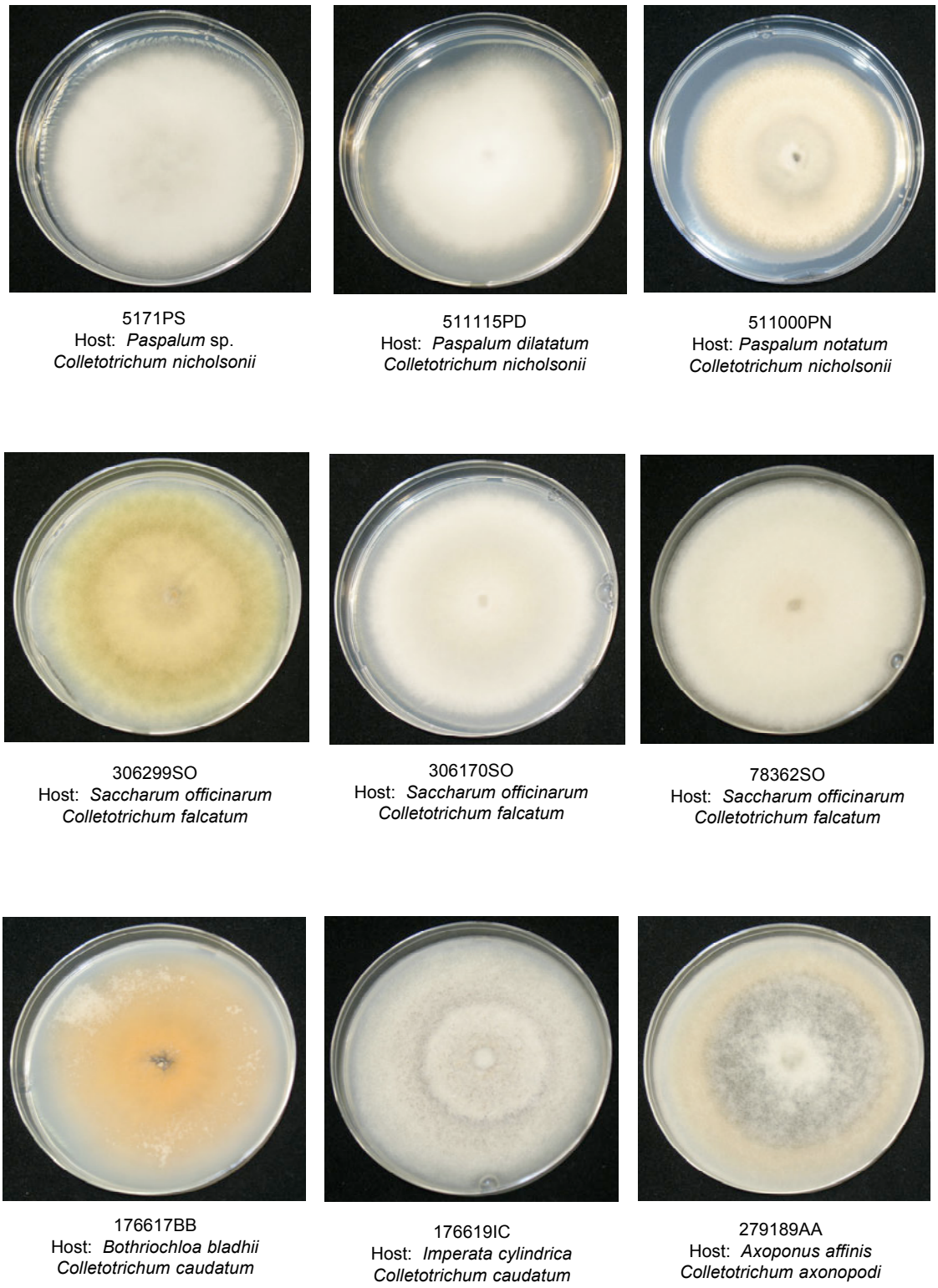


Figure 6.5.1
Isolates of falcate-spored *Colletotrichum* isolated from warm-season grass hosts. Cultures are plated on potato dextrose agar under continuous light, 5-8 d.

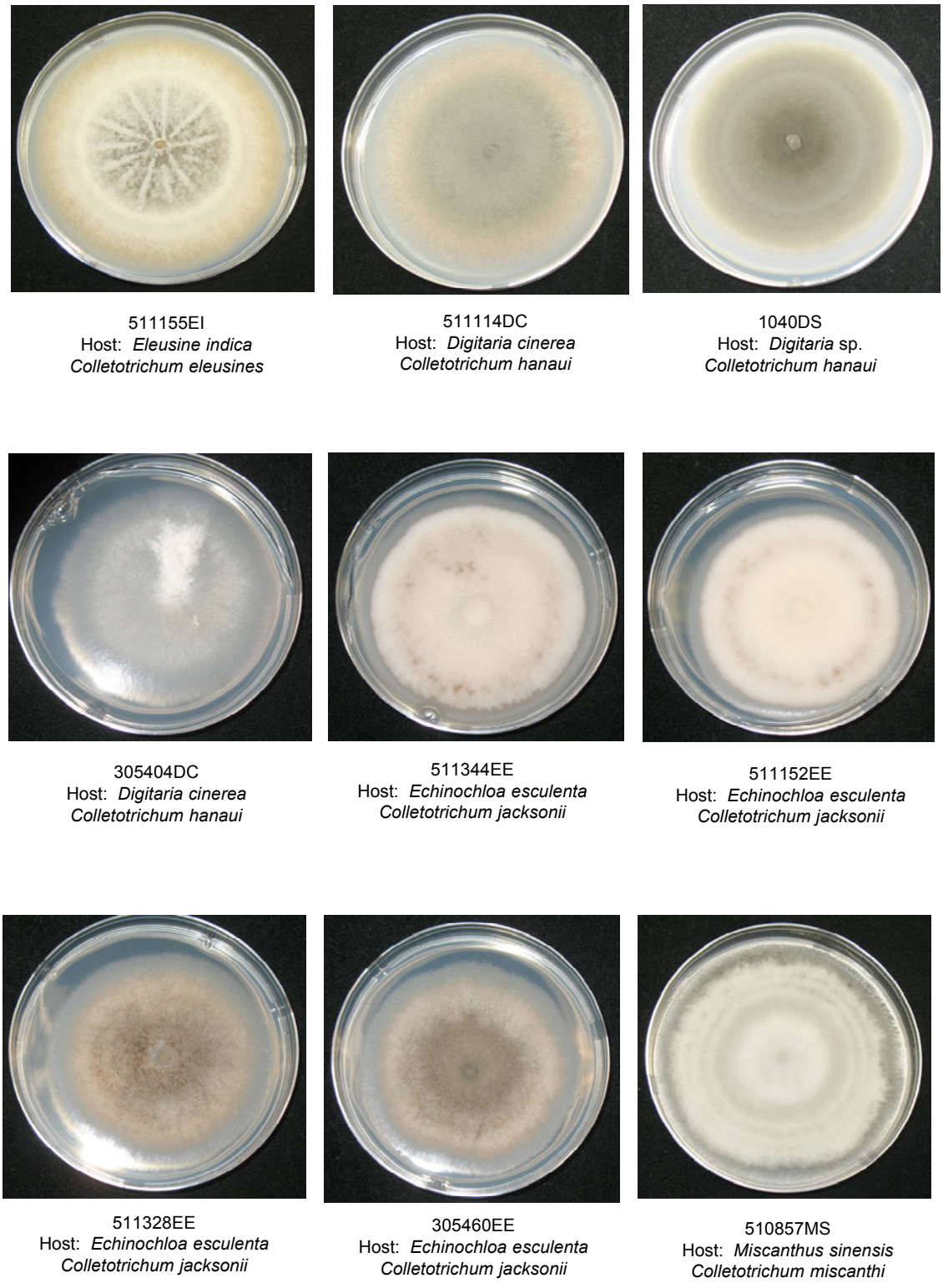
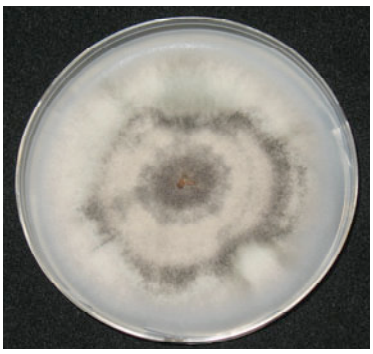
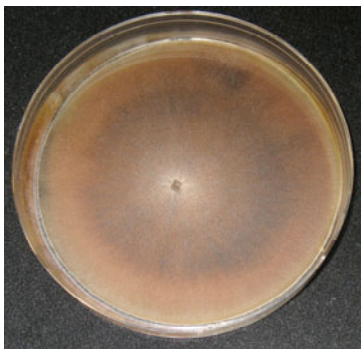


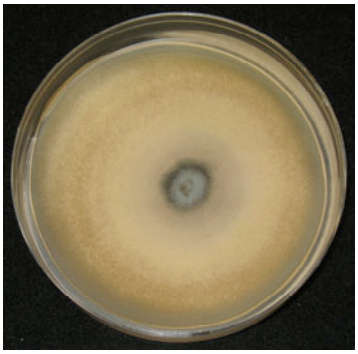
Figure 6.5.2
Isolates of falcate-spored *Colletotrichum* isolated from warm-season grass hosts. Cultures are plated on potato dextrose agar under continuous light, 5-8 d.



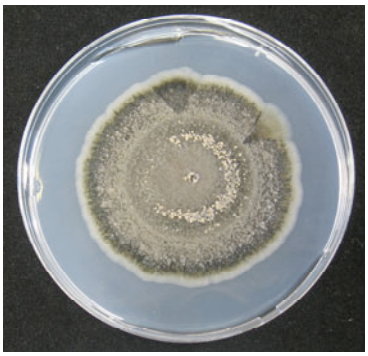
11011031
Host: *Capsicum annuum*
Colletotrichum capsici



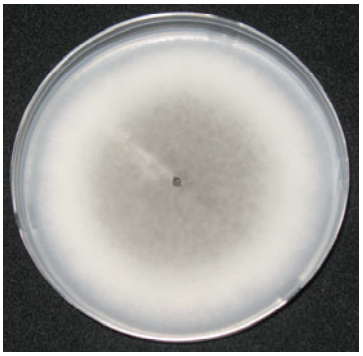
335.75
Host: *Capsicum annuum*
Colletotrichum capsici



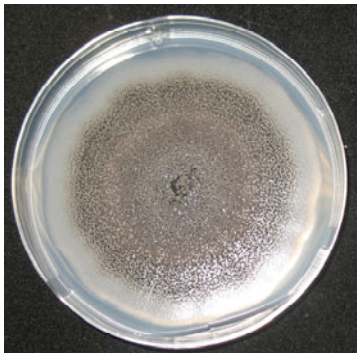
11221
Host: *Linum usitatissimum*
Colletotrichum lini



117546
Host: *Allium porrum*
Colletotrichum circinans

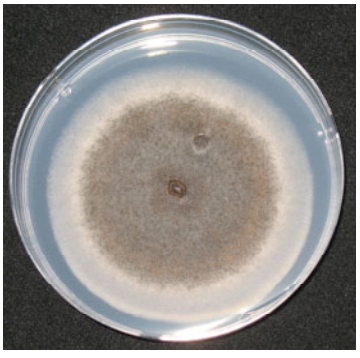


366AV
Host: *Allium vineale*
Colletotrichum circinans

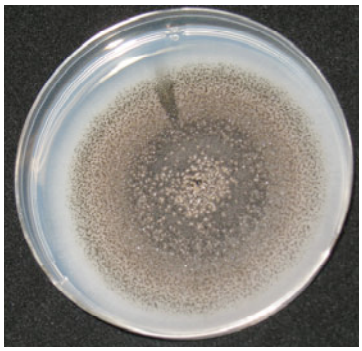


1456GM
Host: *Glycine max*
Colletotrichum truncatum

603DS
Host: *Datura stramonium*
Colletotrichum capsici

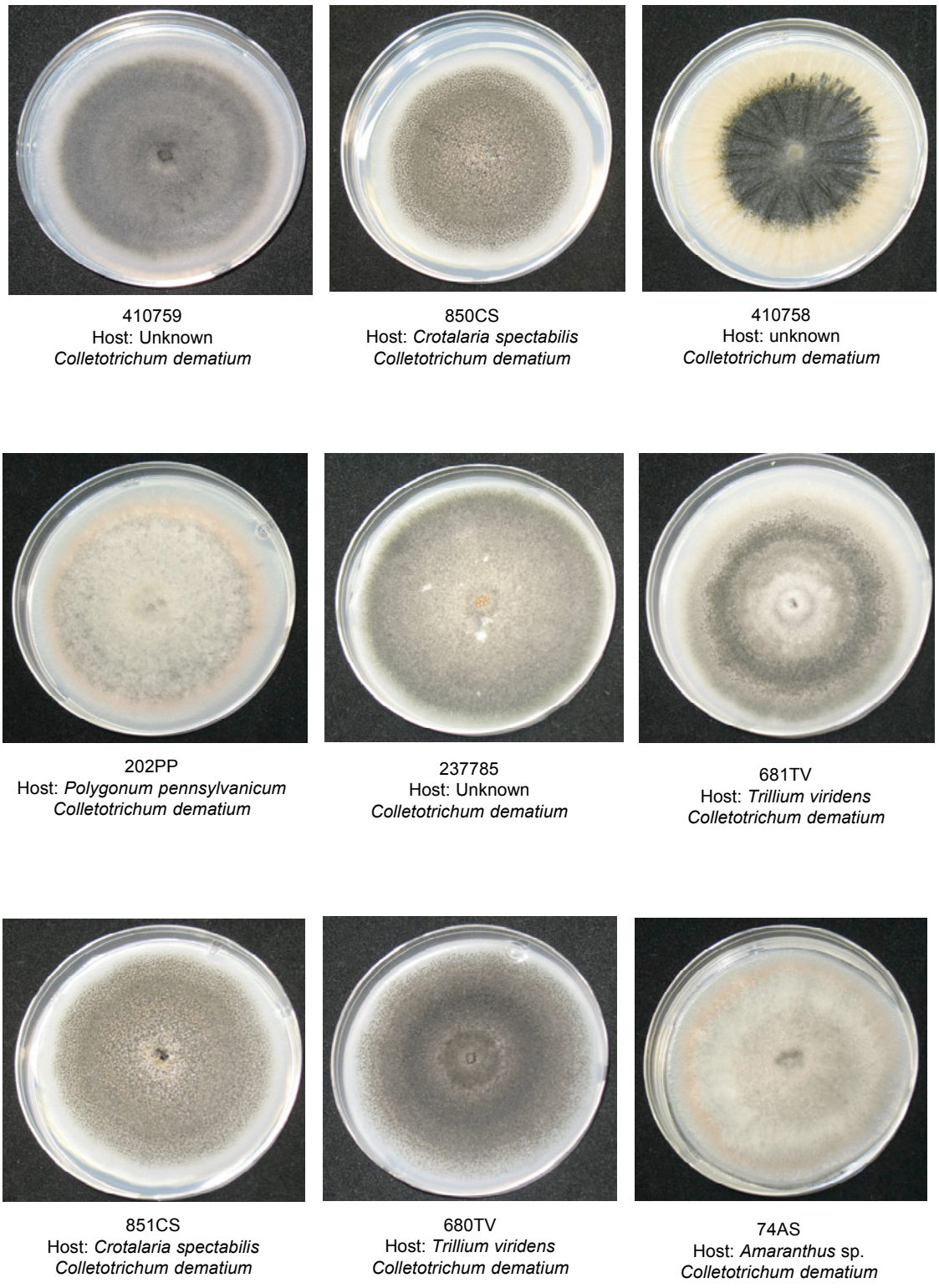


1075PL
Host: *Pueraria lobata*
Colletotrichum dematium



938PL
Host: *Pueraria lobata*
Colletotrichum dematium

Figure 6.6.1
Isolates of falcate-spored *Colletotrichum* isolated from non-graminicolous hosts. Cultures are plated on potato dextrose agar under continuous light, 5-8 d.



410759
Host: Unknown
Colletotrichum dematium

850CS
Host: *Crotalaria spectabilis*
Colletotrichum dematium

410758
Host: unknown
Colletotrichum dematium

202PP
Host: *Polygonum pennsylvanicum*
Colletotrichum dematium

237785
Host: Unknown
Colletotrichum dematium

681TV
Host: *Trillium viridens*
Colletotrichum dematium

851CS
Host: *Crotalaria spectabilis*
Colletotrichum dematium

680TV
Host: *Trillium viridens*
Colletotrichum dematium

74AS
Host: *Amaranthus* sp.
Colletotrichum dematium

Figure 6.6.2
Isolates of falcate-spored *Colletotrichum* isolated from non-graminicolous hosts. Cultures are plated on potato dextrose agar under continuous light, 5-8 d.

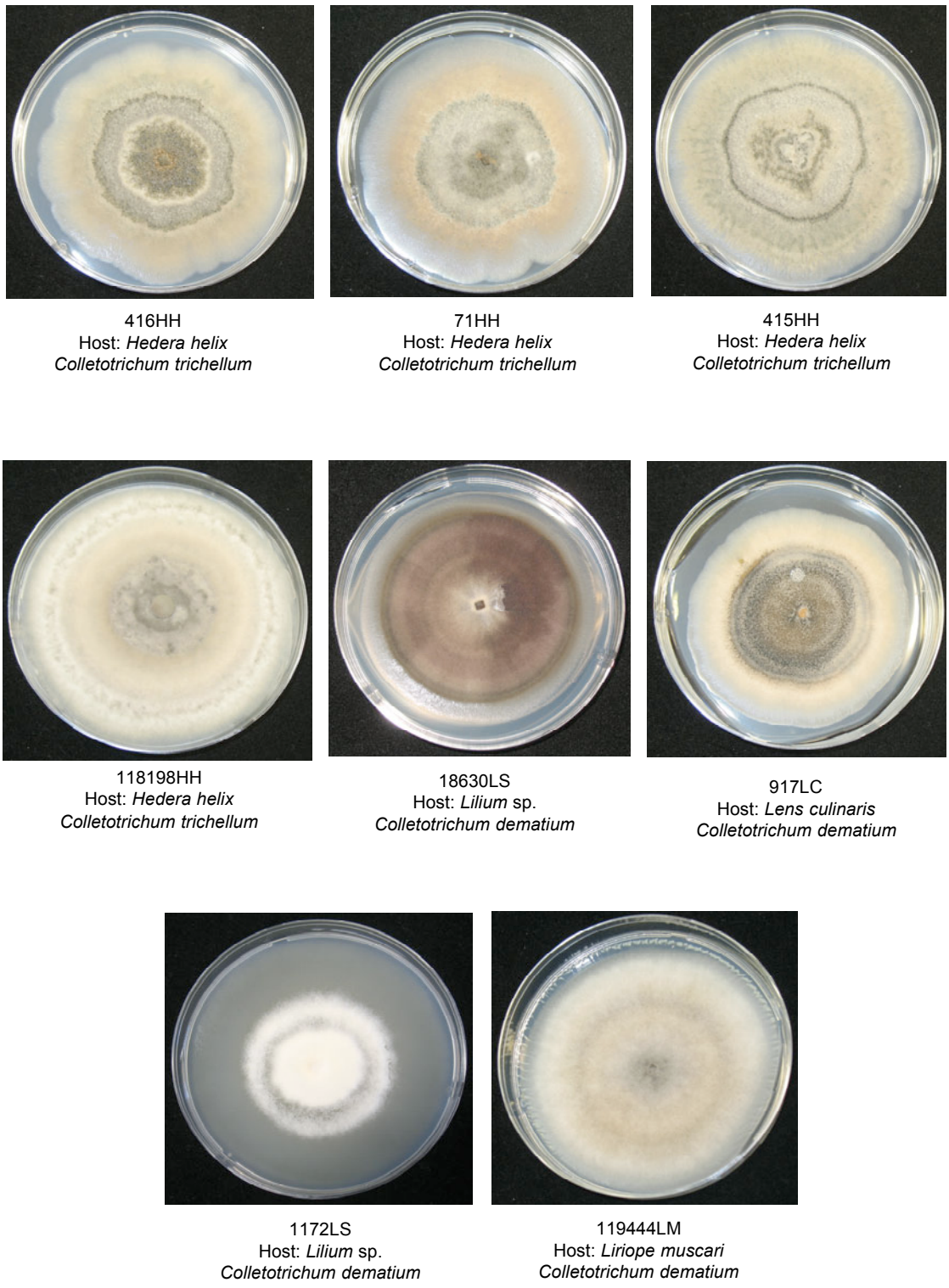
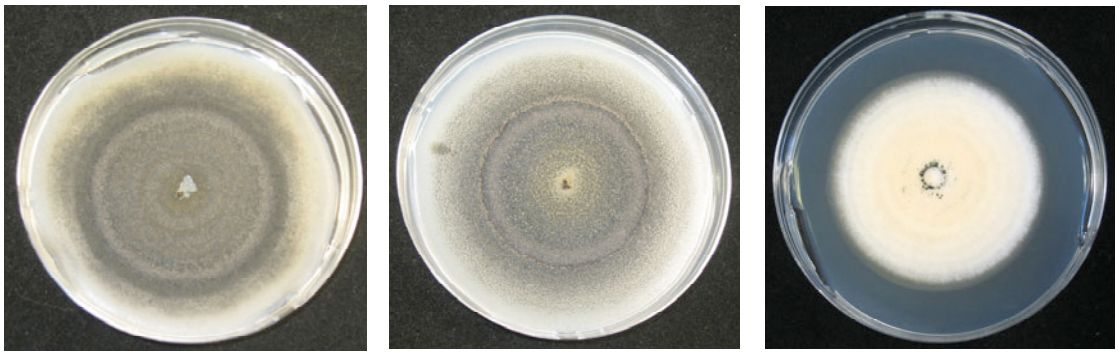


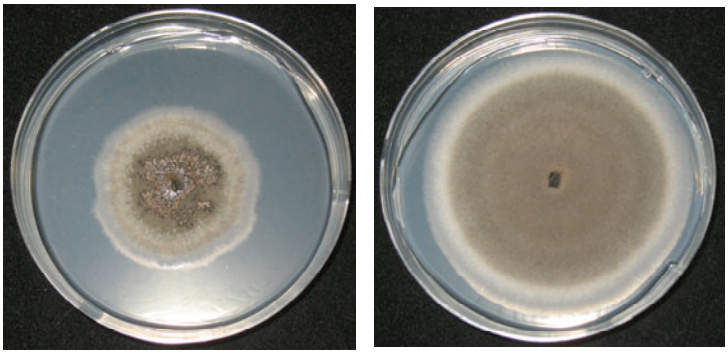
Figure 6.6.3
Isolates of falcate-spored *Colletotrichum* isolated from non-graminicolous hosts. Cultures are plated on potato dextrose agar under continuous light, 5-8 d.



81067
Host: Unknown
Colletotrichum dematium

84066
Host: Unknown
Colletotrichum dematium

11AA
Host: *Aeschynomene americana*
Colletotrichum dematium



IRI-5 (A)
Colletotrichum gloeosporioides

IRI-5 (B)
Colletotrichum gloeosporioides

Figure 6.6.4
Pictures 1-3 (upper row): Isolates of falcate-spored *Colletotrichum* isolated from non-graminicolous hosts. Pictures 4-5 (lower row): isolates of the oval-spored *C. gloeosporioide*s from palm leaves Cultures are plated on potato dextrose agar under continuous light, 5-8 d.

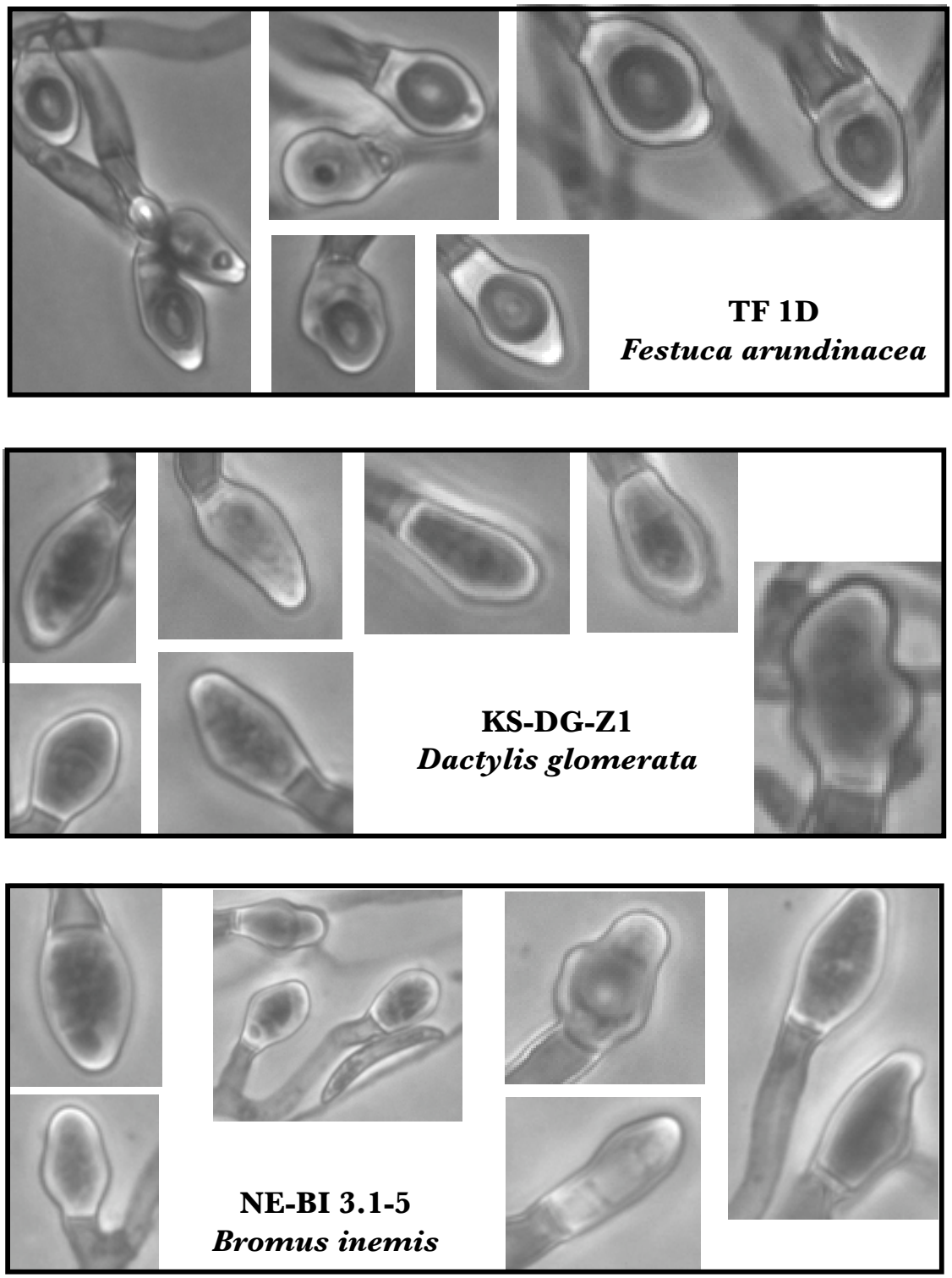


Figure 6.7.1
Hyphopodia of *Colletotrichum cereale*. Bar=25 μm.

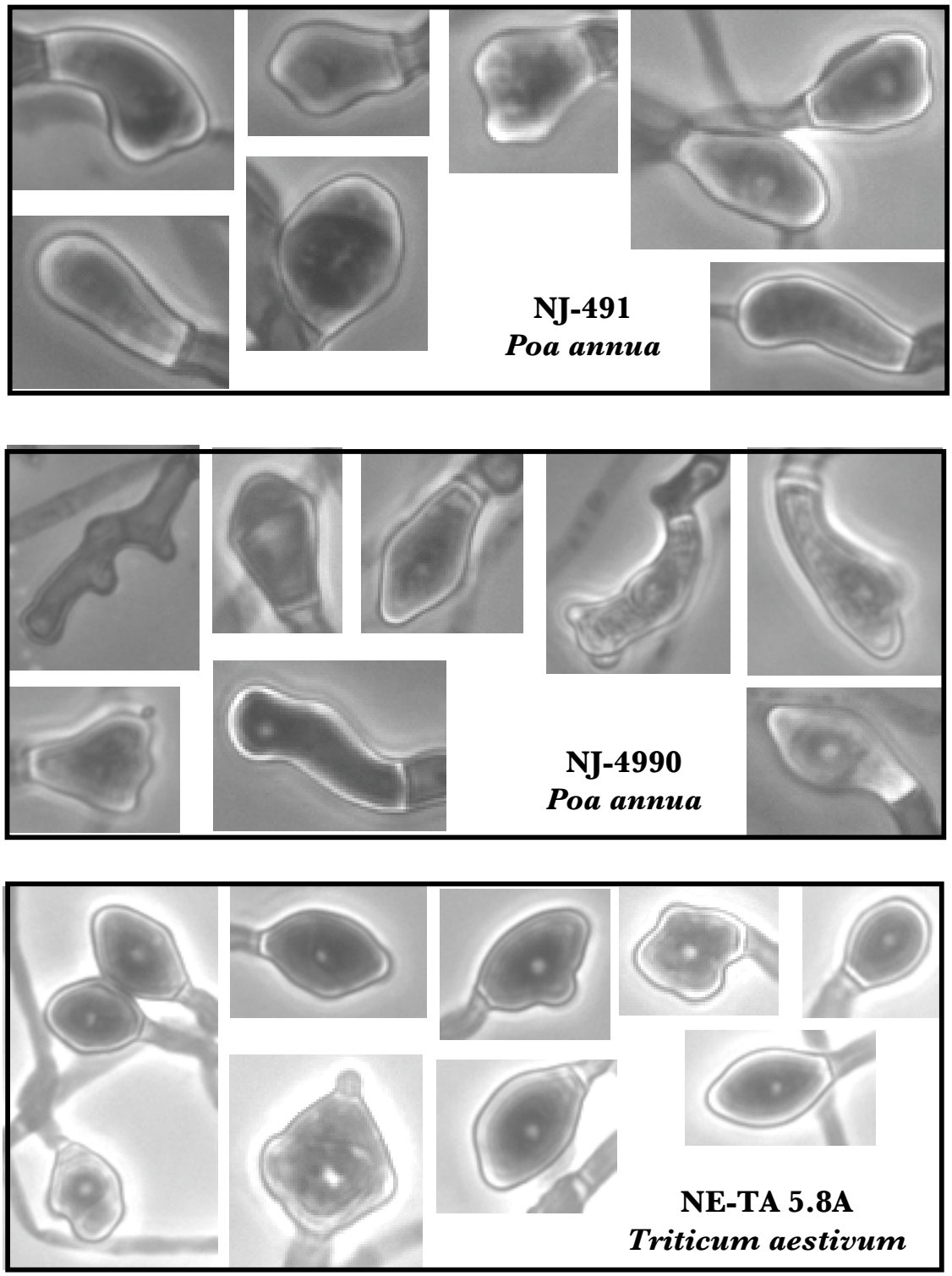


Figure 6.7.2
Hyphopodia of *Colletotrichum cereale*. Bar=25 μm.

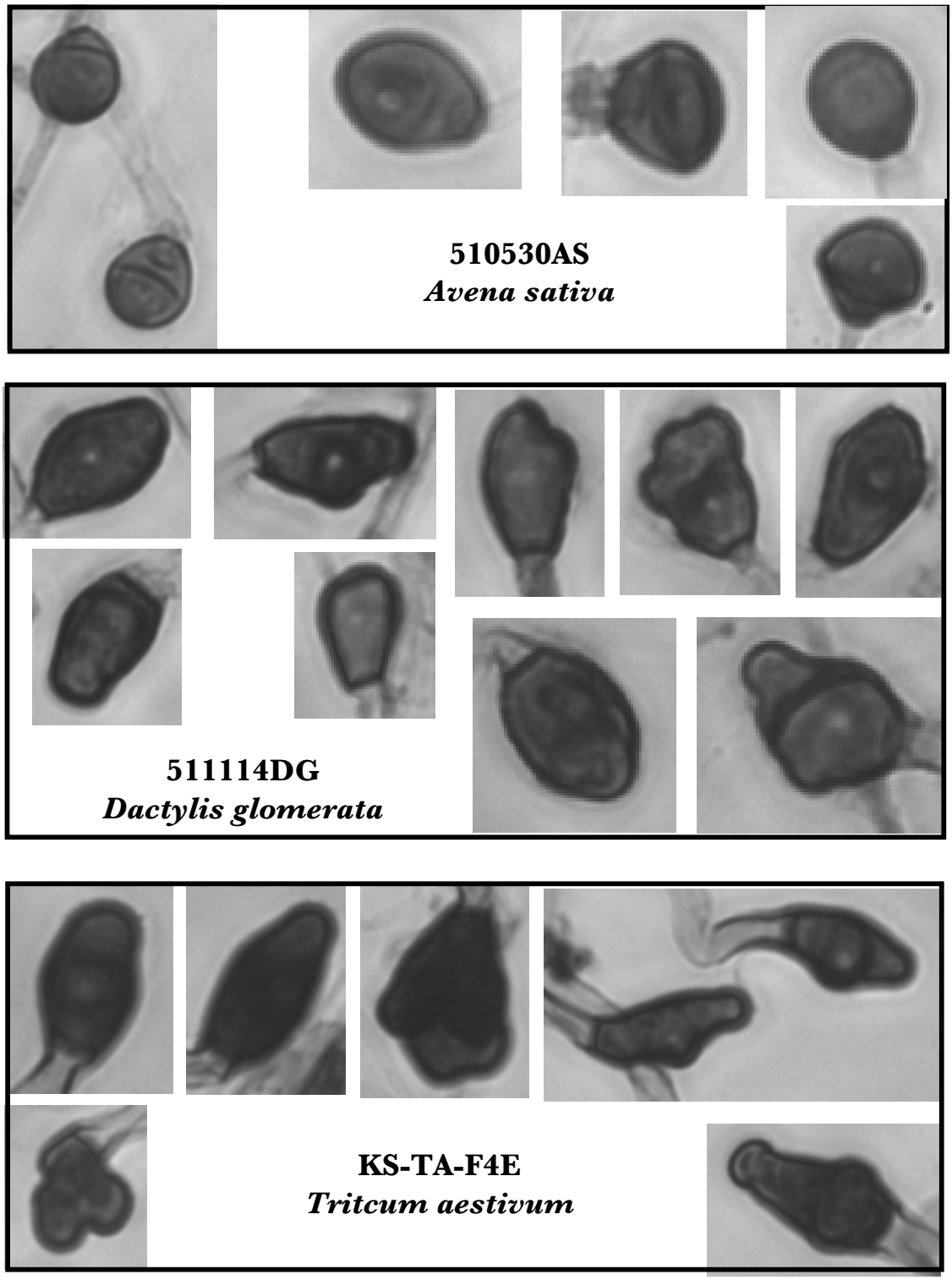


Figure 6.7.4
Hyphopodia of *Colletotrichum cereale*. Bar=25 μm.

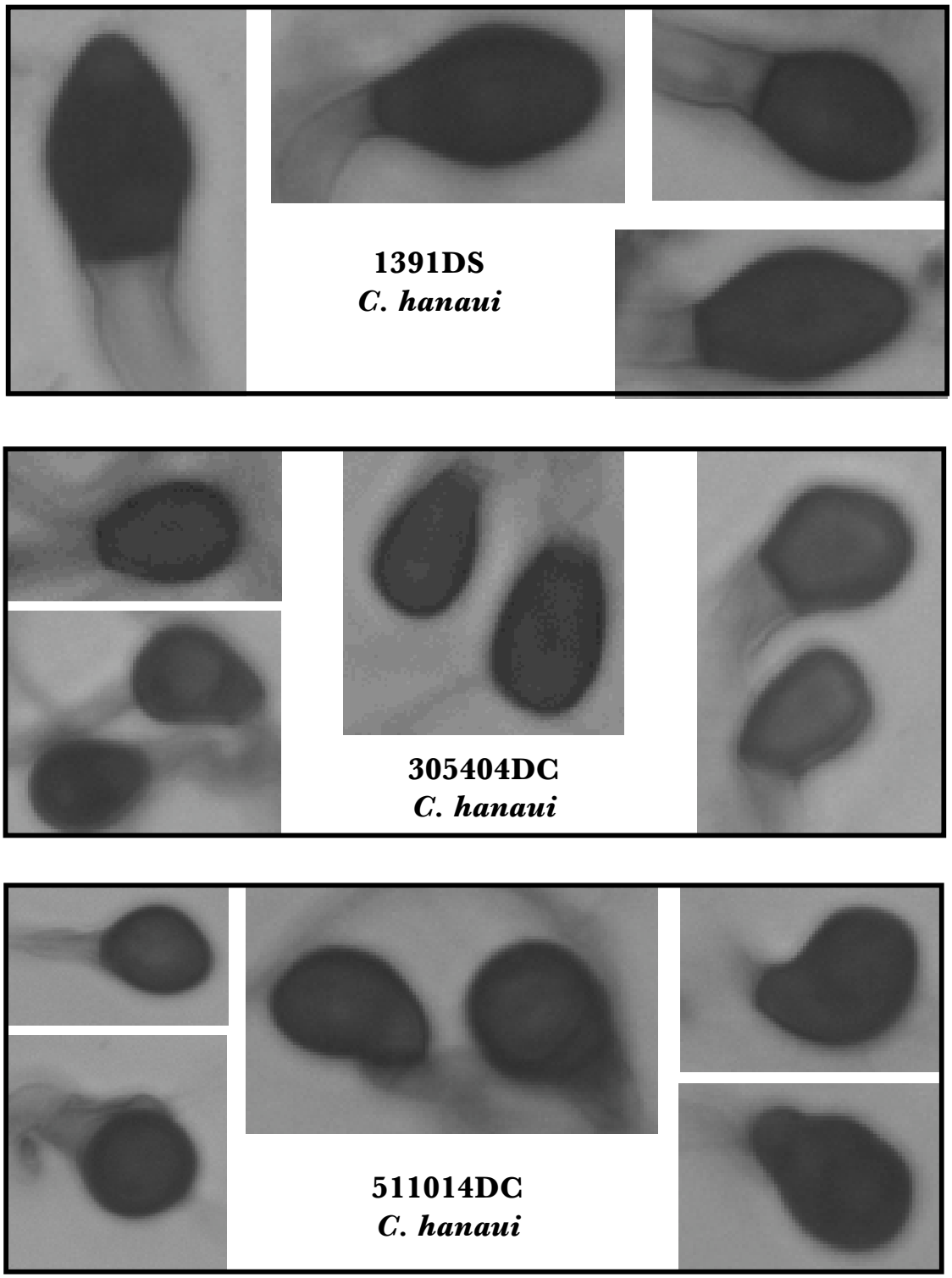


Figure 6.8.1
Hyphopodia of *Colletotrichum hanaui* isolated from warm season grasses of the genus *Digitaria*. Bar=25 μm.

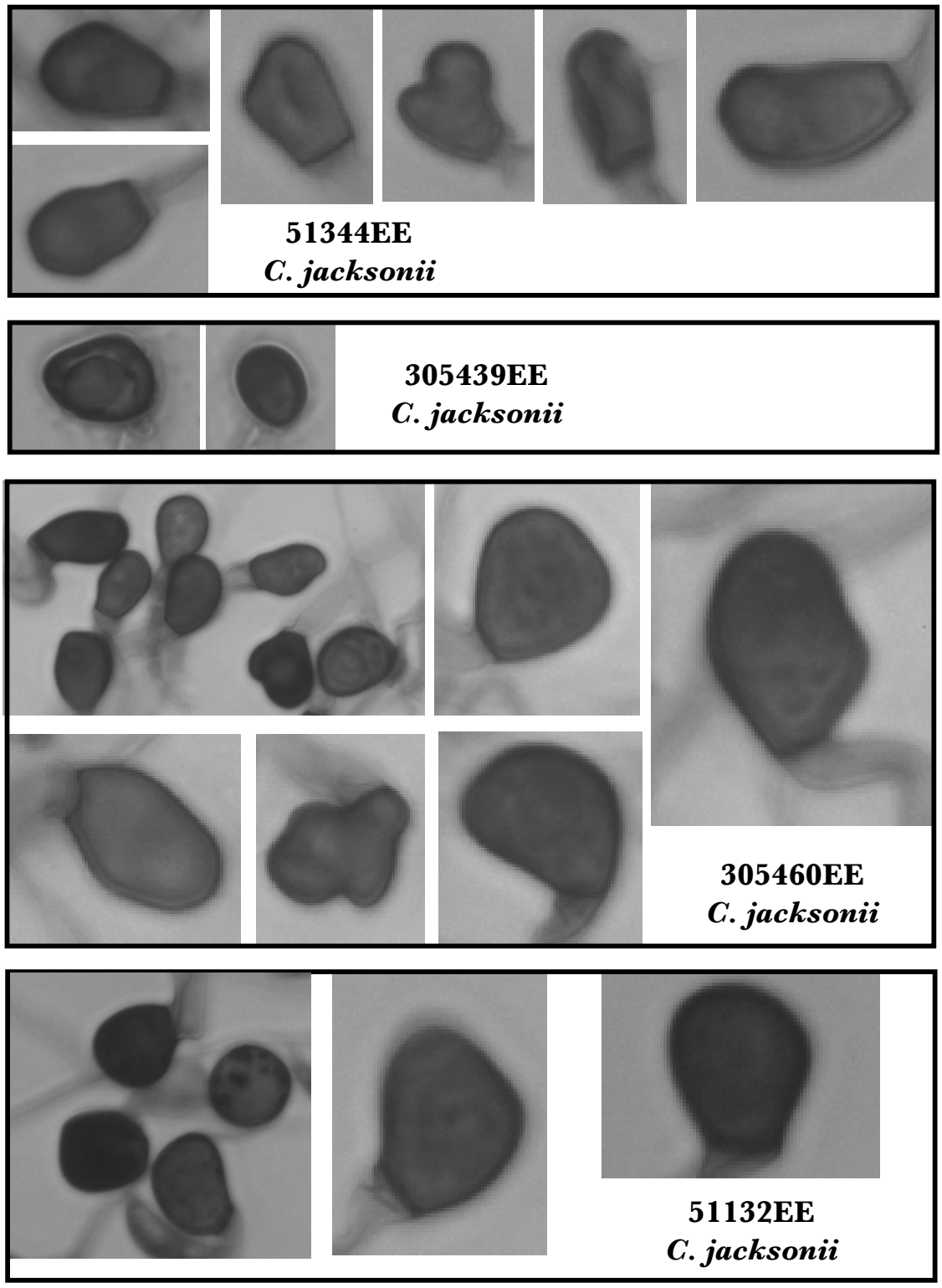


Figure 6.8.2
Hyphopodia of *Colletotrichum jacksonii* isolated from the warm season grass *Echinochloa esculenta*. Bar=25 μm.

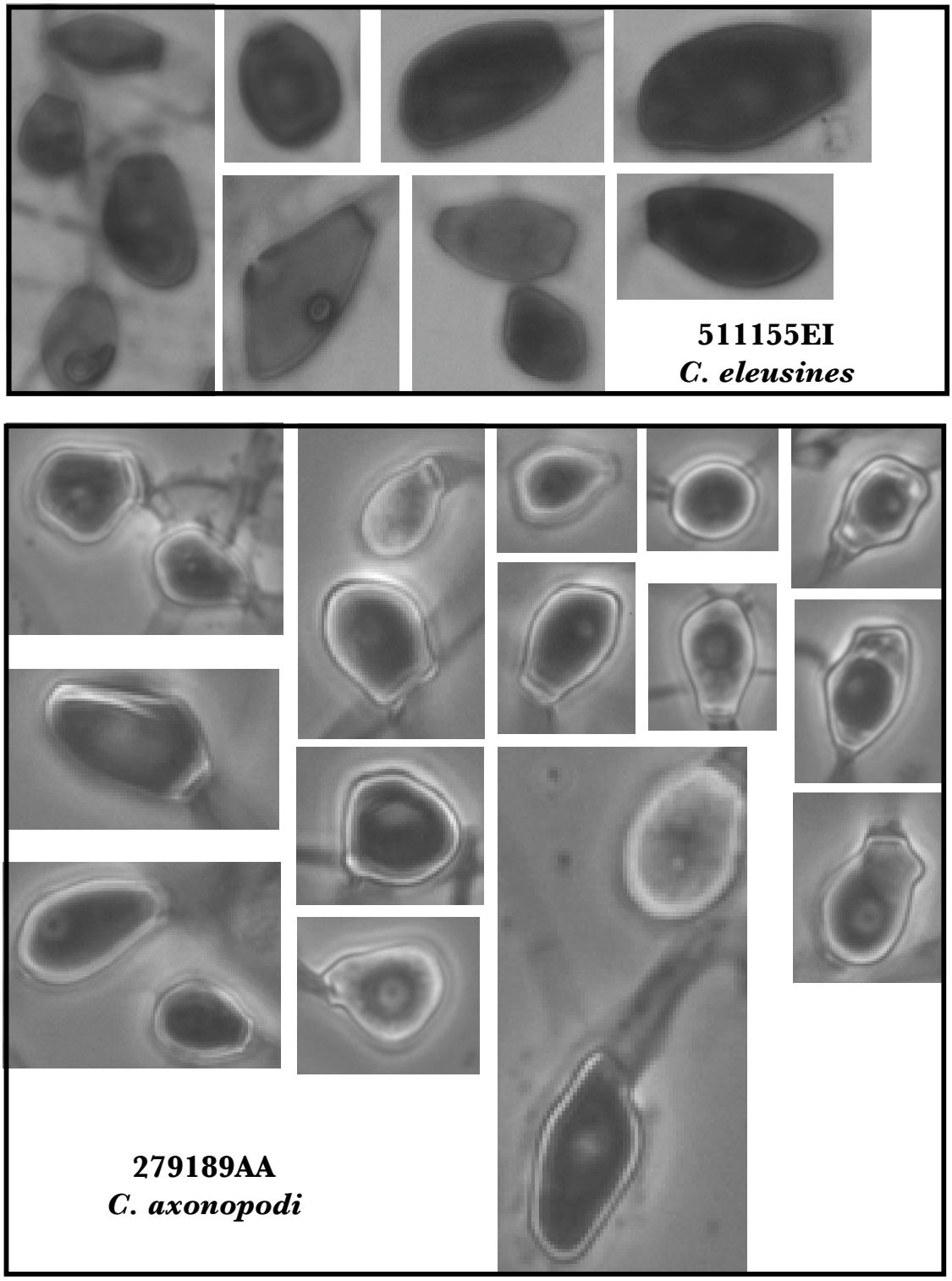


Figure 6.8.3
Hyphopodia of *Colletotrichum eleusines* and *C. axonopodi* isolated from warm season grasses of the genus *Eleusine indica* and *Axonopus affinis*. Bar=25 μm.

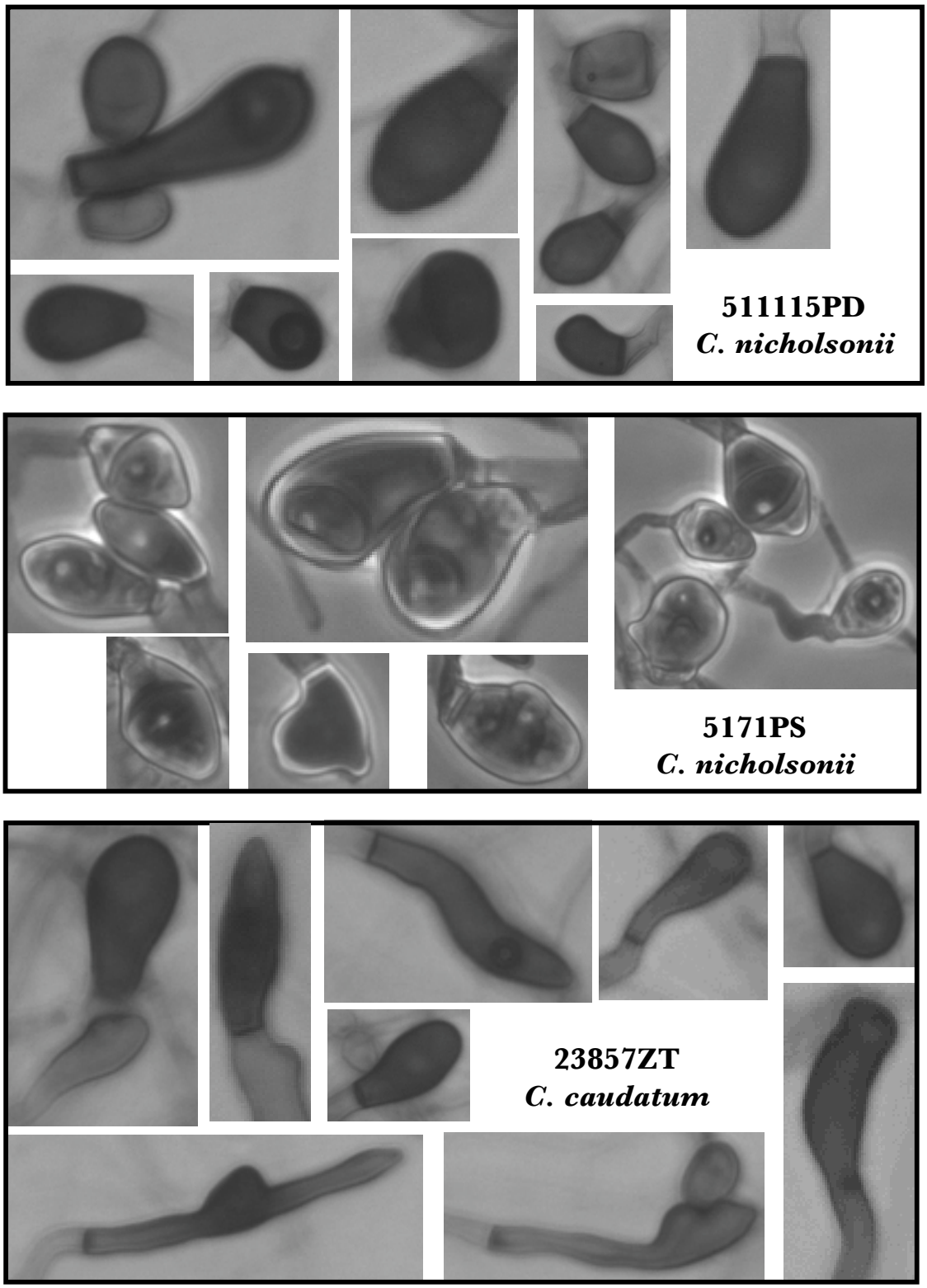


Figure 6.8.4
Hyphopodia of *Colletotrichum nicholsonii* and *C. caudatum* isolated from warm season grasses of the genus *Paspalum* and *Zoysia*. Bar=25 μm.

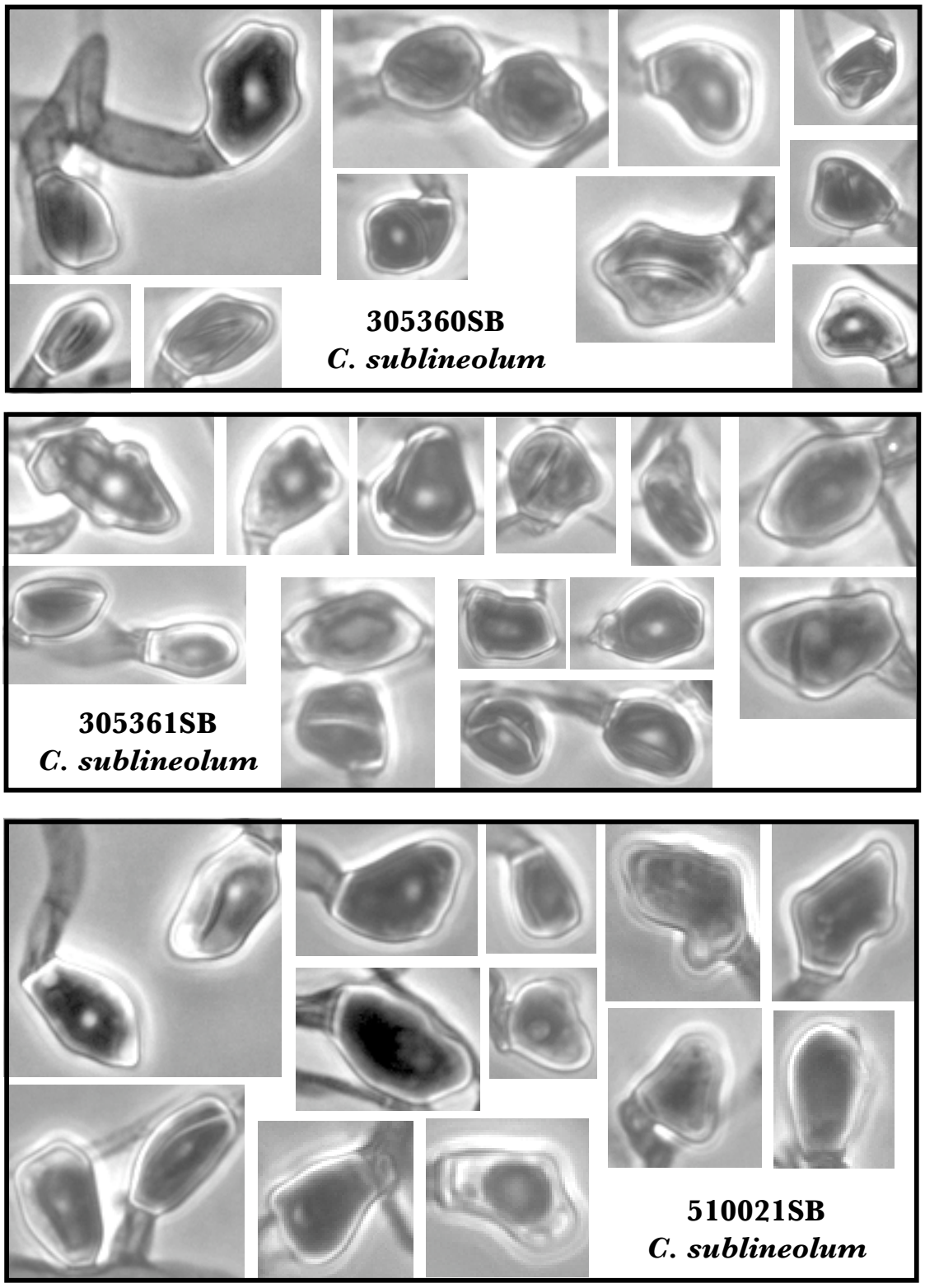


Figure 6.8.5
Hyphopodia of *Colletotrichum sublineolum* isolated from warm season grasses of the genus *Sorghum*. Bar=25 μm.

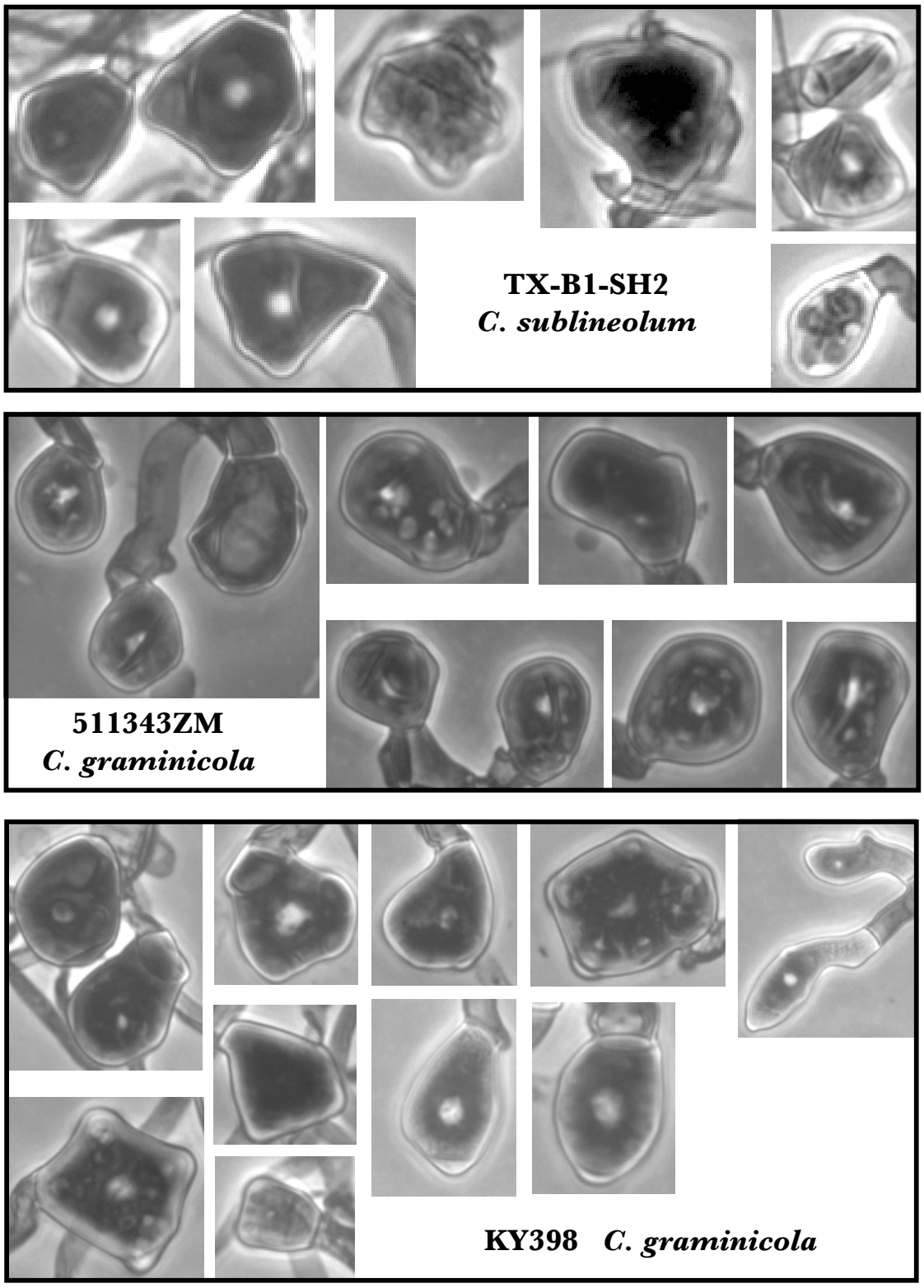


Figure 6.8.6
Hyphopodia of *Colletotrichum sublineolum* and *C. graminicola* isolated from warm season grasses of the genus *Sorghum* and *Zea*. Bar=25 μm.

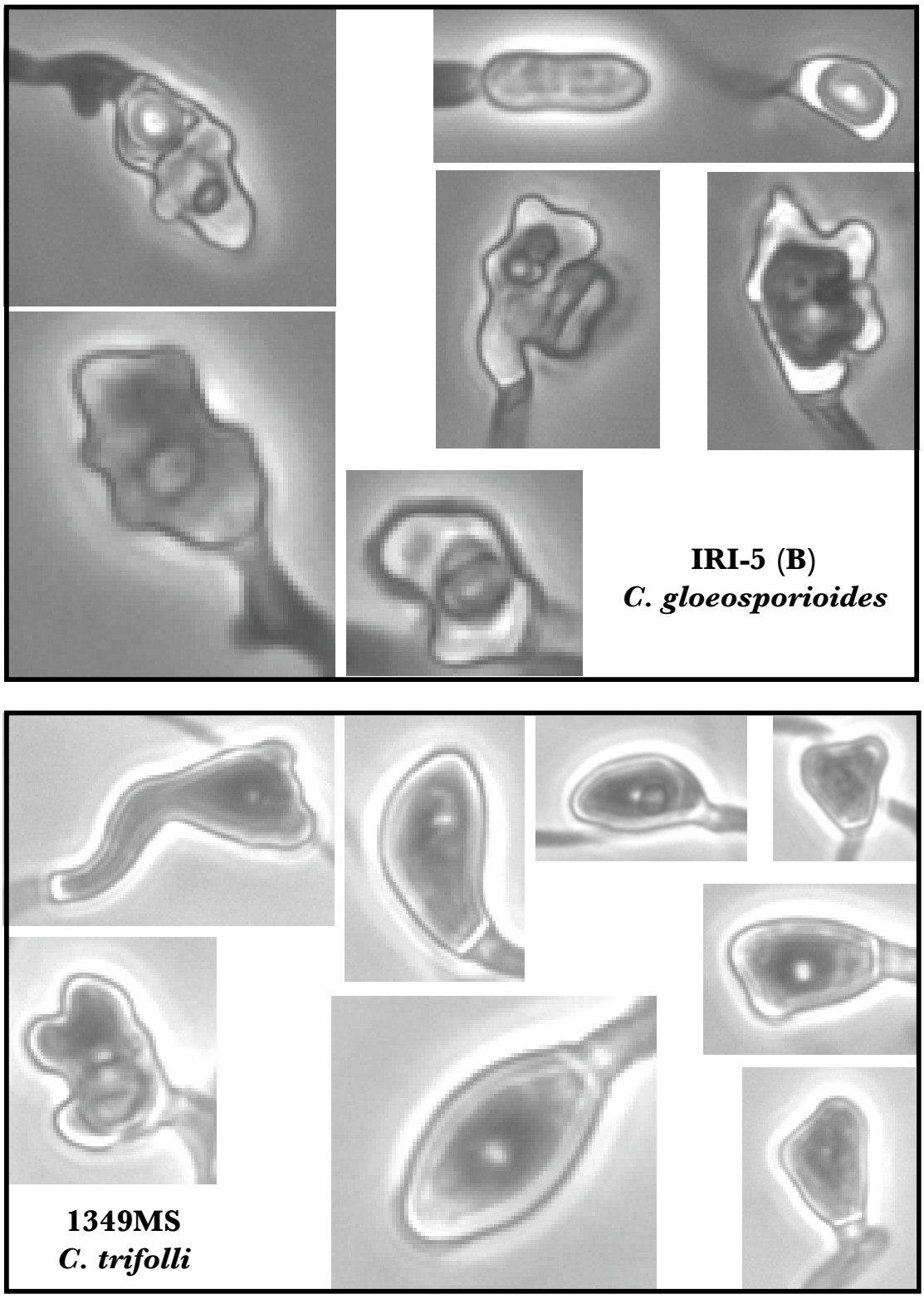


Figure 6.9.1
Hyphopodia of *Colletotrichum* isolated non-graminicolous plant hosts. Bar=25 μm.

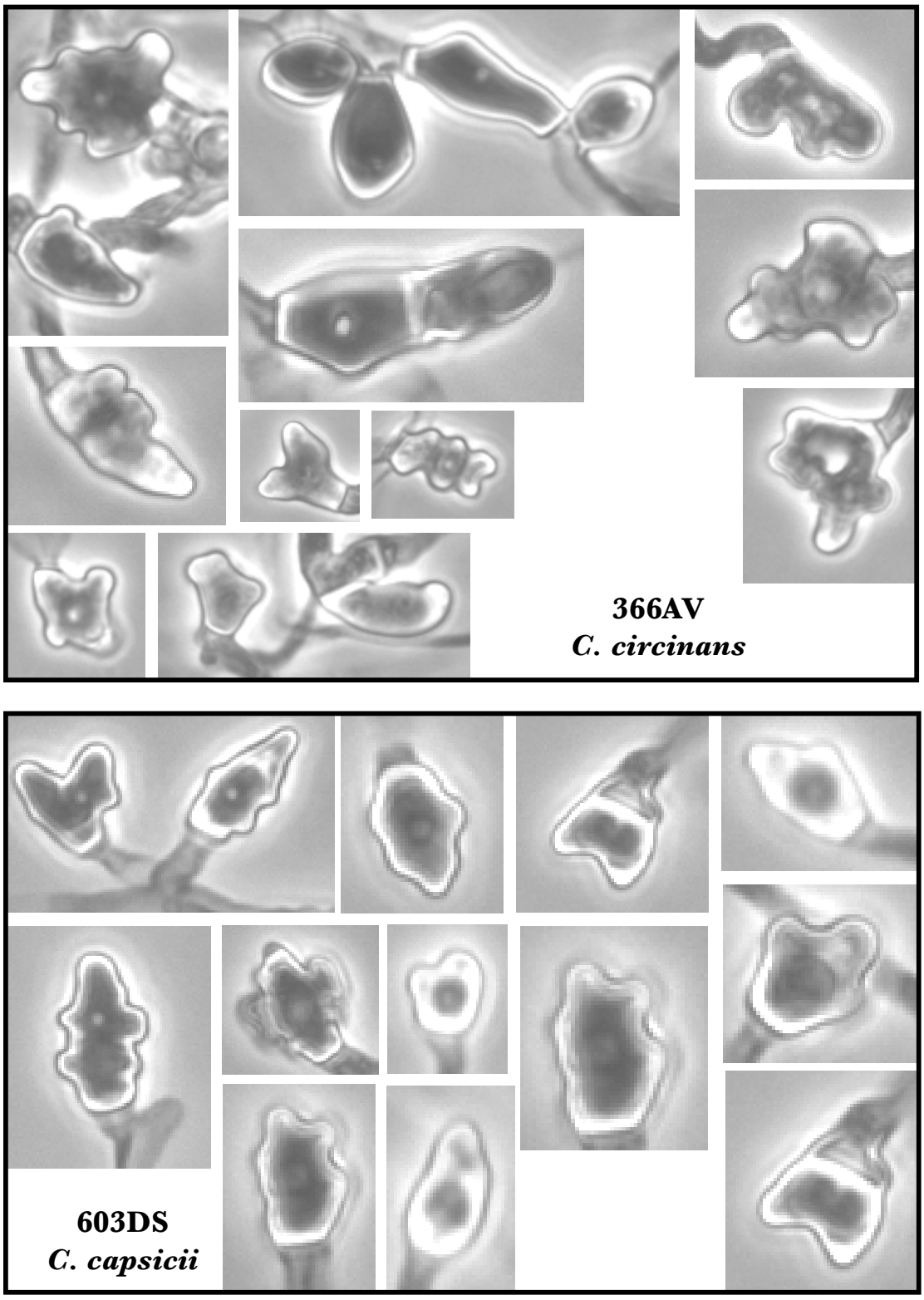


Figure 6.9.2
Hyphopodia of *Colletotrichum* isolated non-graminicolous plant hosts. Bar=25 μm.

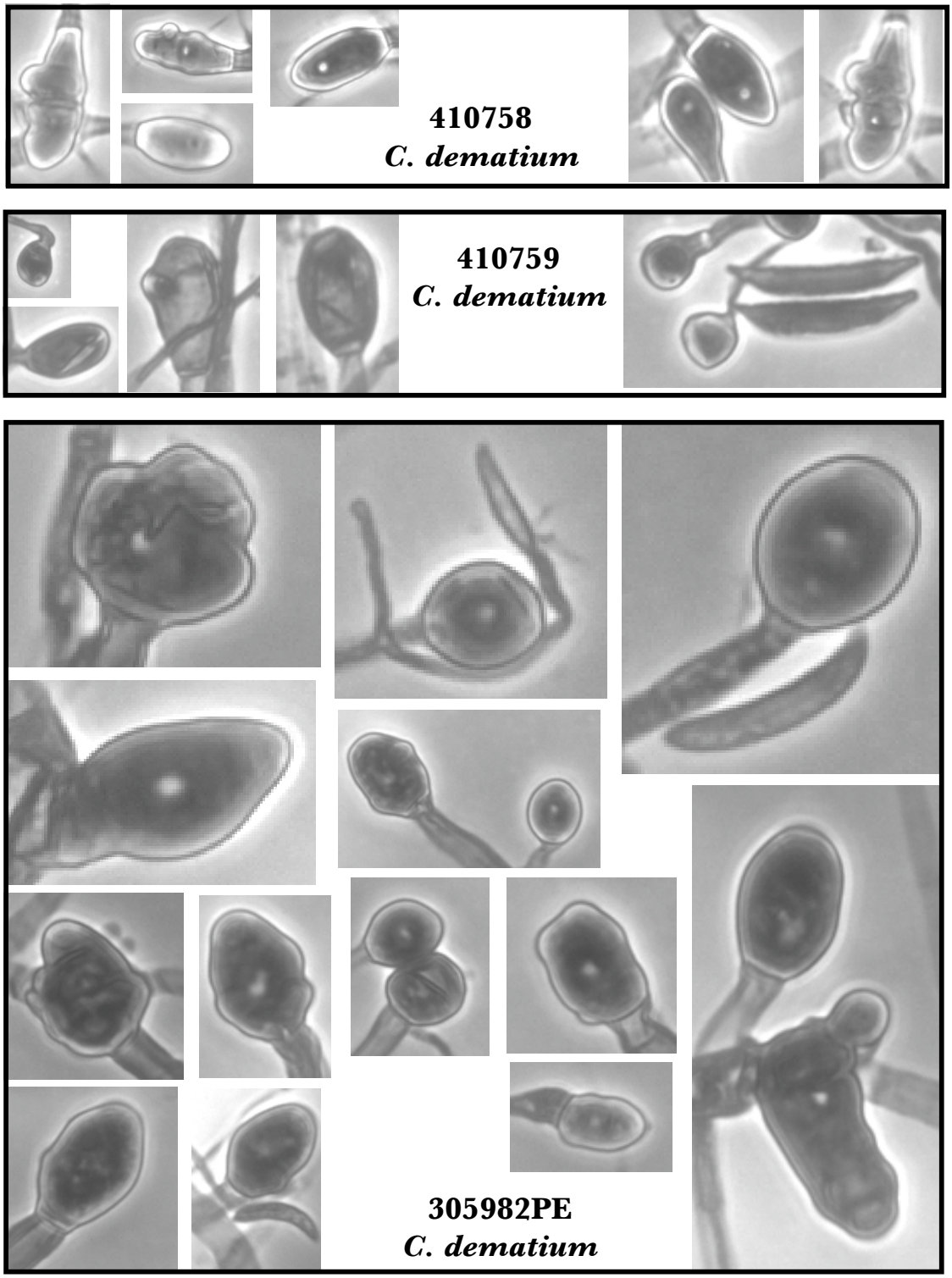


Figure 6.9.3
Hyphopodia of *Colletotrichum* isolated from non-graminicolous hosts. Bar=25 μm.

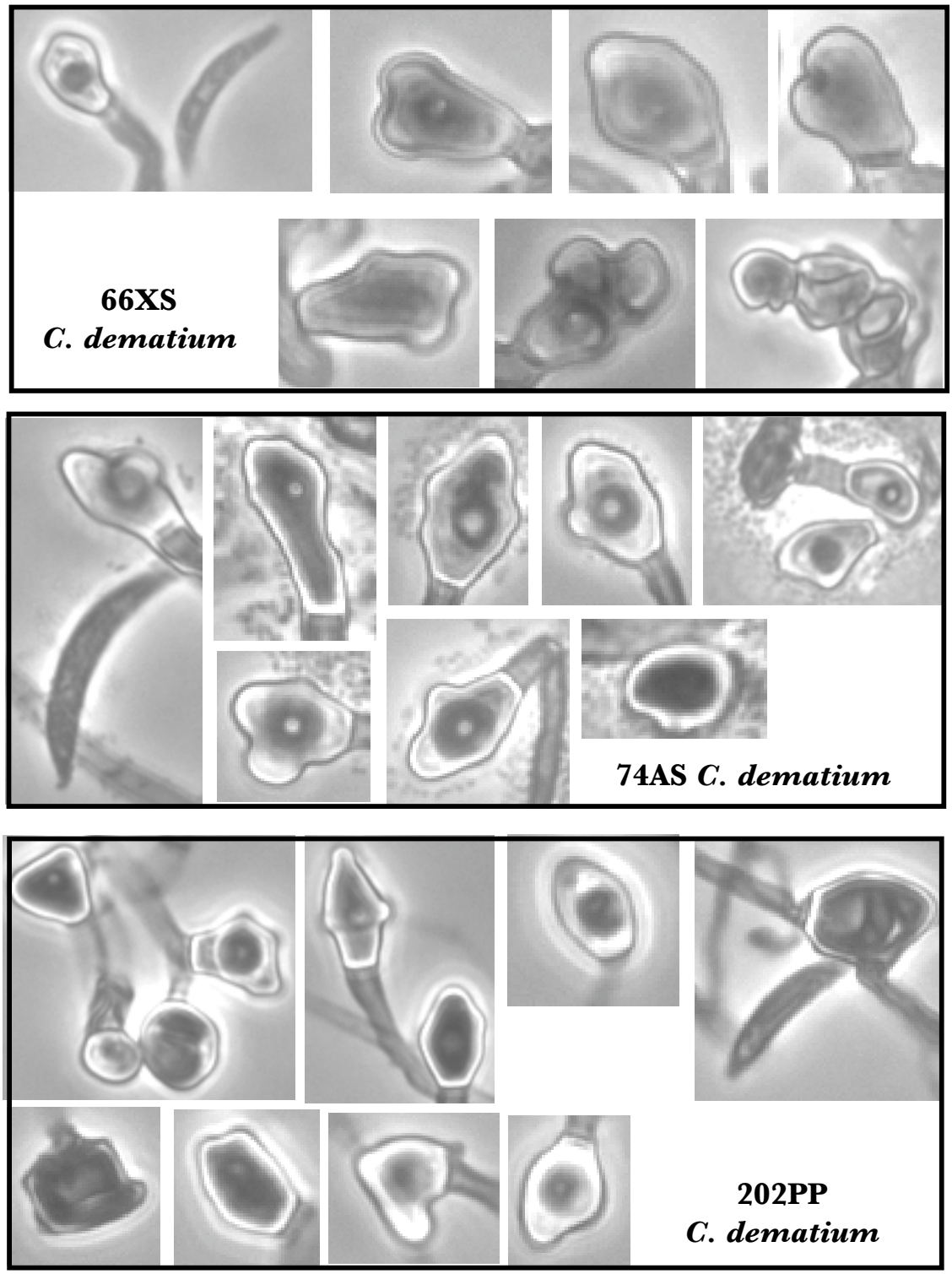


Figure 6.9.4
Hyphopodia of *Colletotrichum* isolated from non-graminicolous hosts. Bar=25 μm.

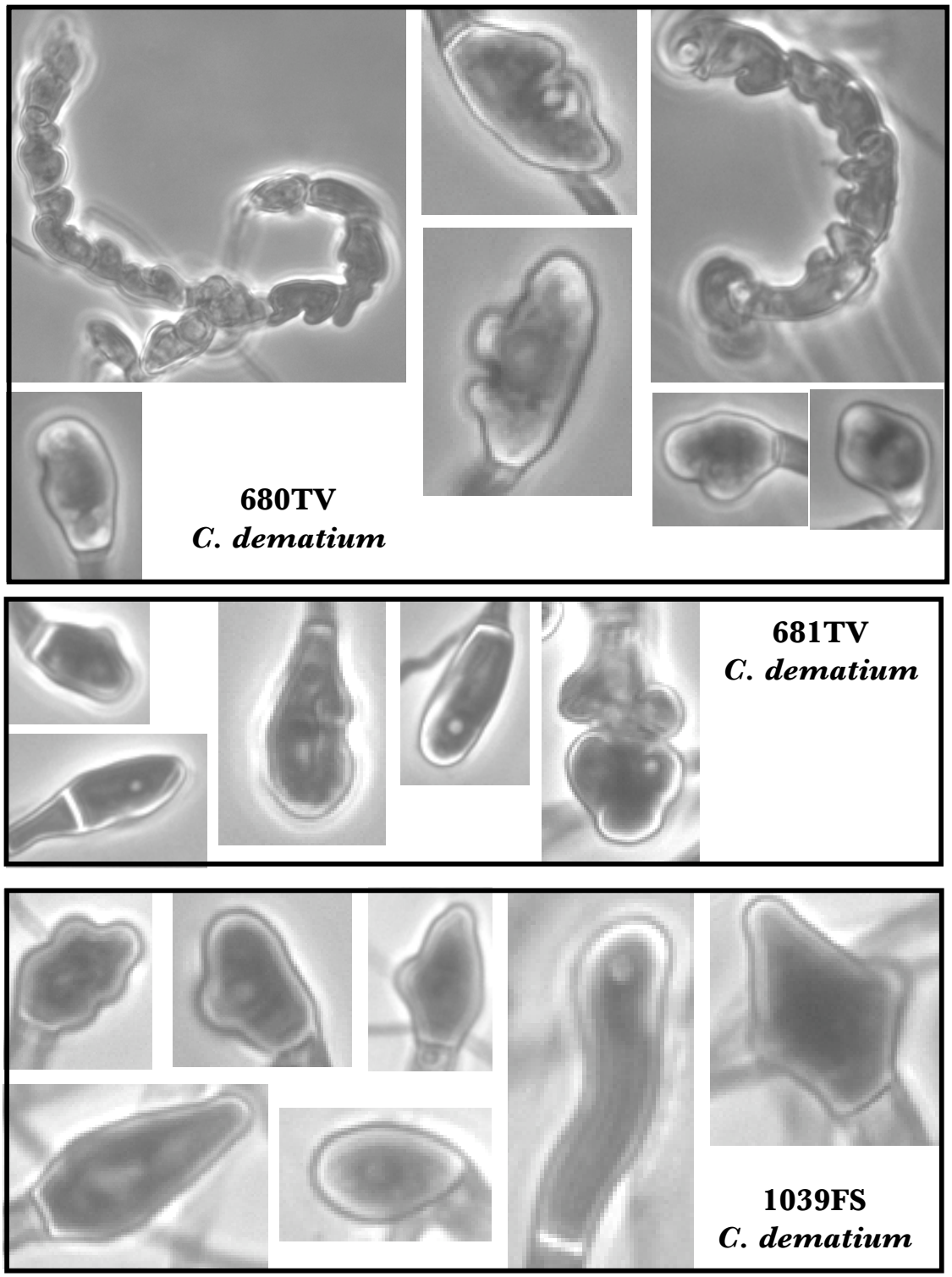


Figure 6.9.5
Hyphopodia of *Colletotrichum* isolated from non-graminicolous hosts. Bar=25 μm.

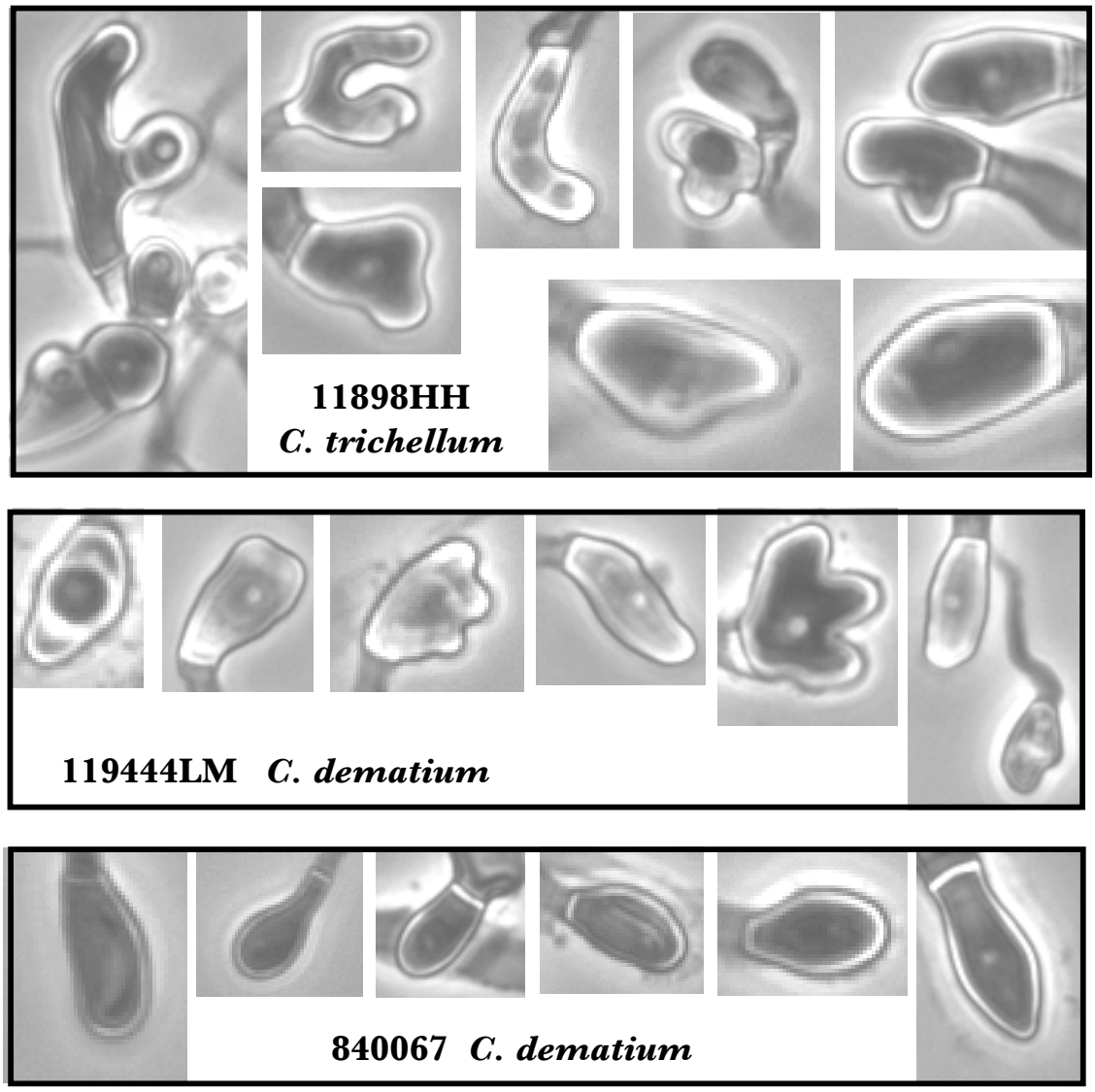


Figure 6.9.6
Hyphopodia of *Colletotrichum* isolated from non-graminicolous hosts. Bar=25 μm.

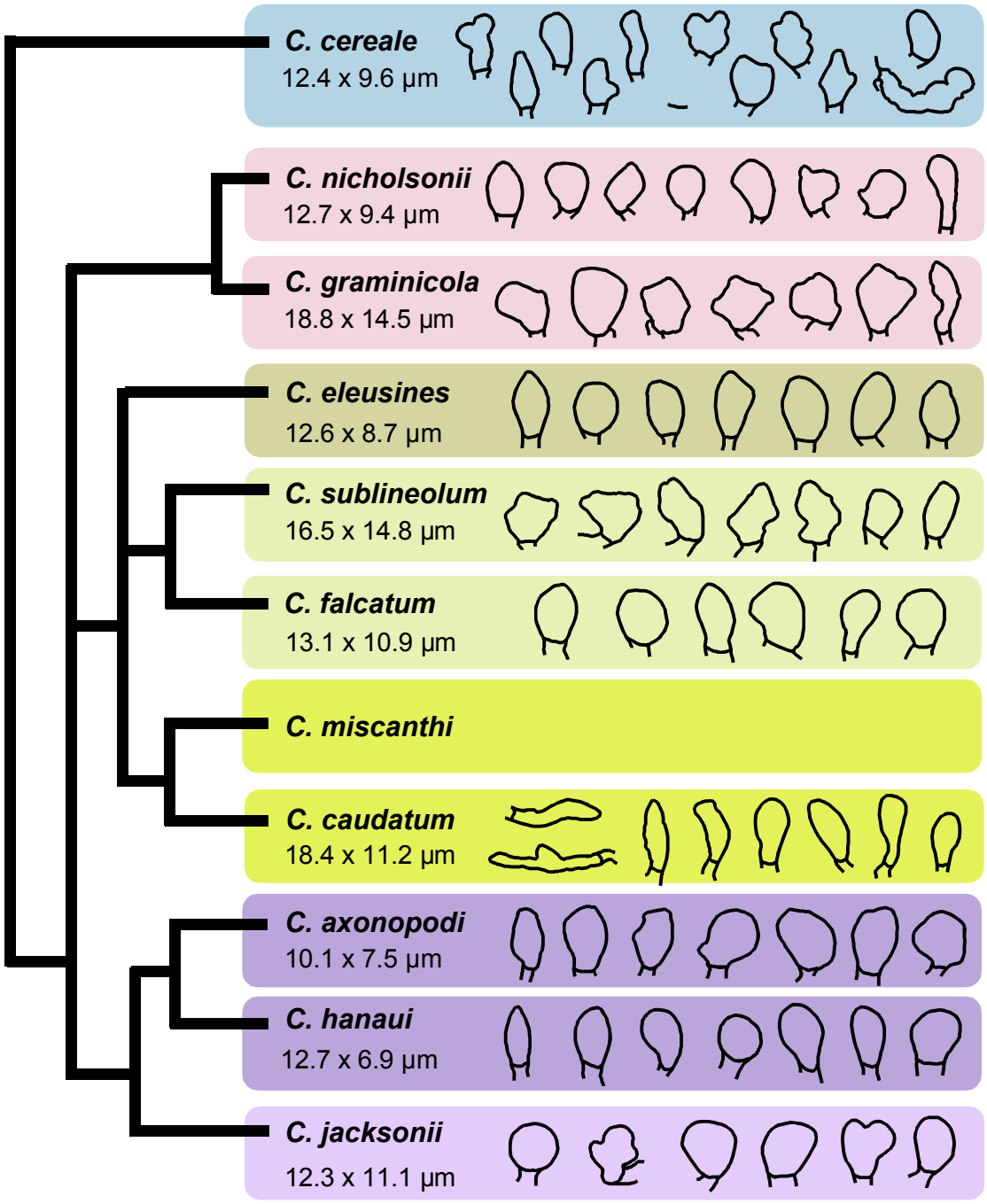


Figure 6.10
Evolutionary groups within the graminicolous *Colletotrichum* as inferred through molecular phylogenetic analysis. Hyphopodial appressoria size (avergae height x width; Table 3) and shape are plotted along the phylogeny. Illustrations of hyphopodia are not to scale; they are meant to illustrate the range of shapes observed.

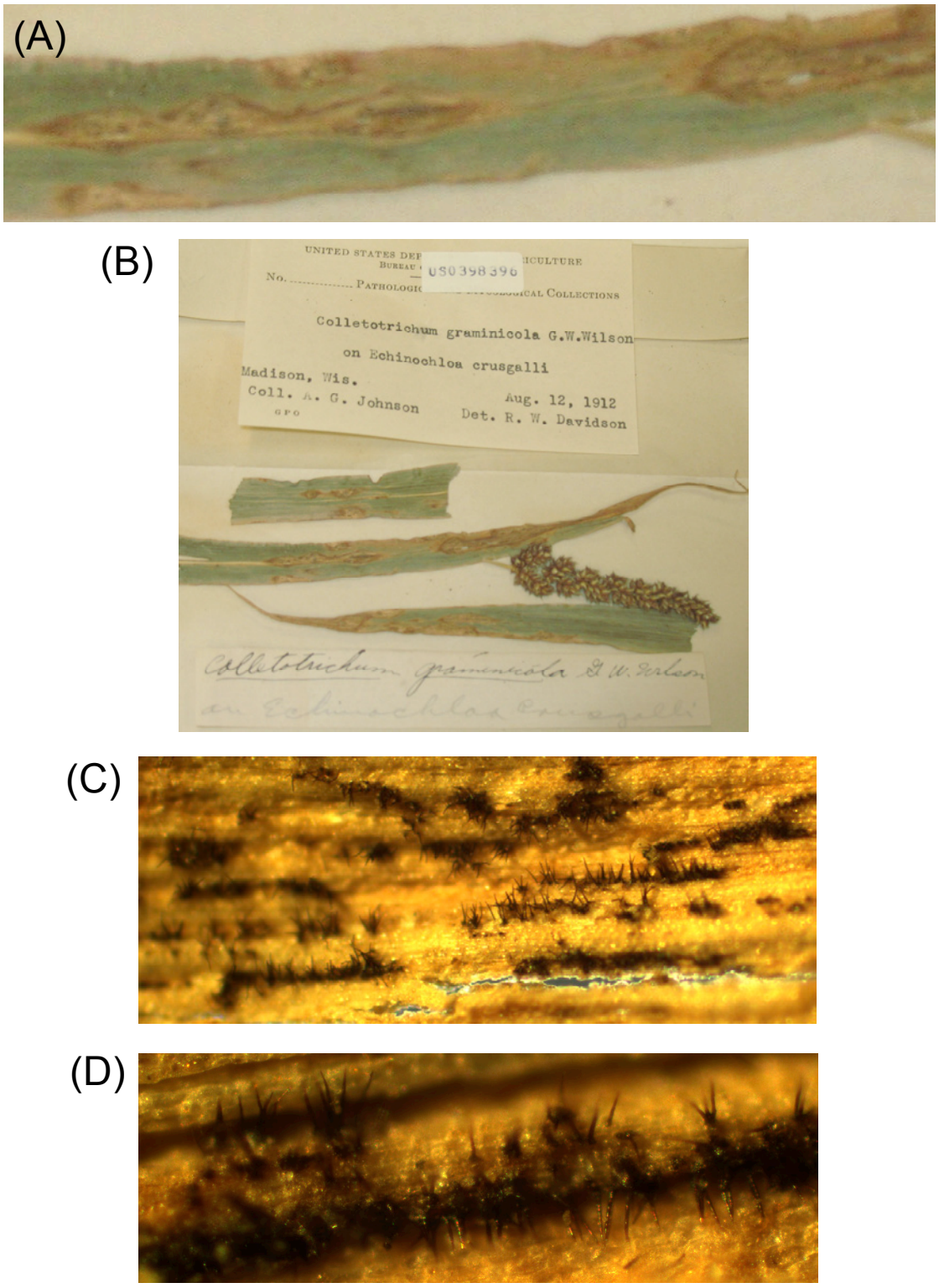
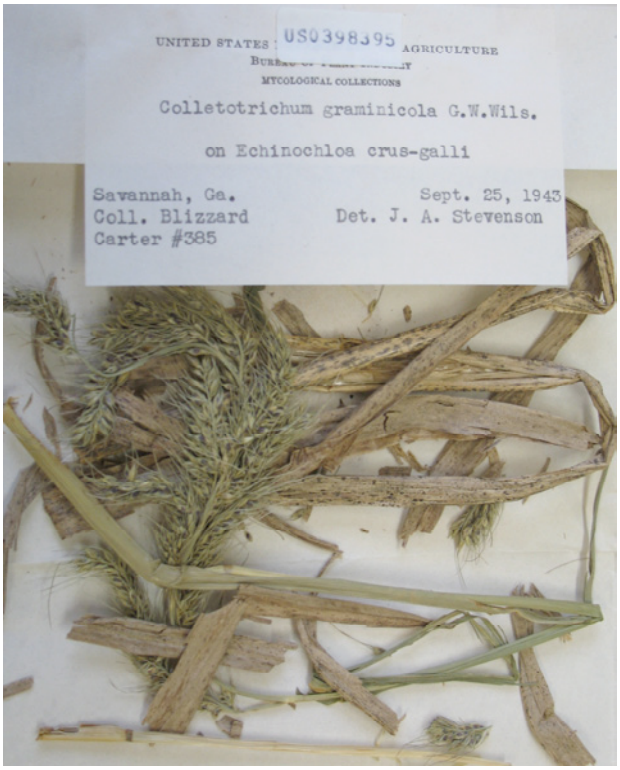


Figure 6.11.1
Herbarium specimens of *C. jacksonii*. (A) and (B) disease symptoms on infected *Echinochloa crus-galli*, (C) and (D) *C. jacksonii* setae, 400x magification. BPI398396, collected 12-Aug-1912.

(A)



(B)

(C)

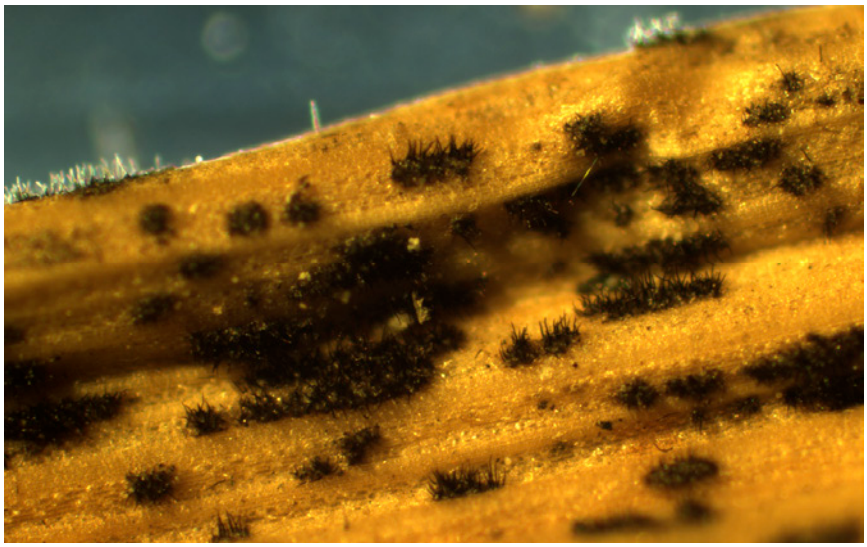
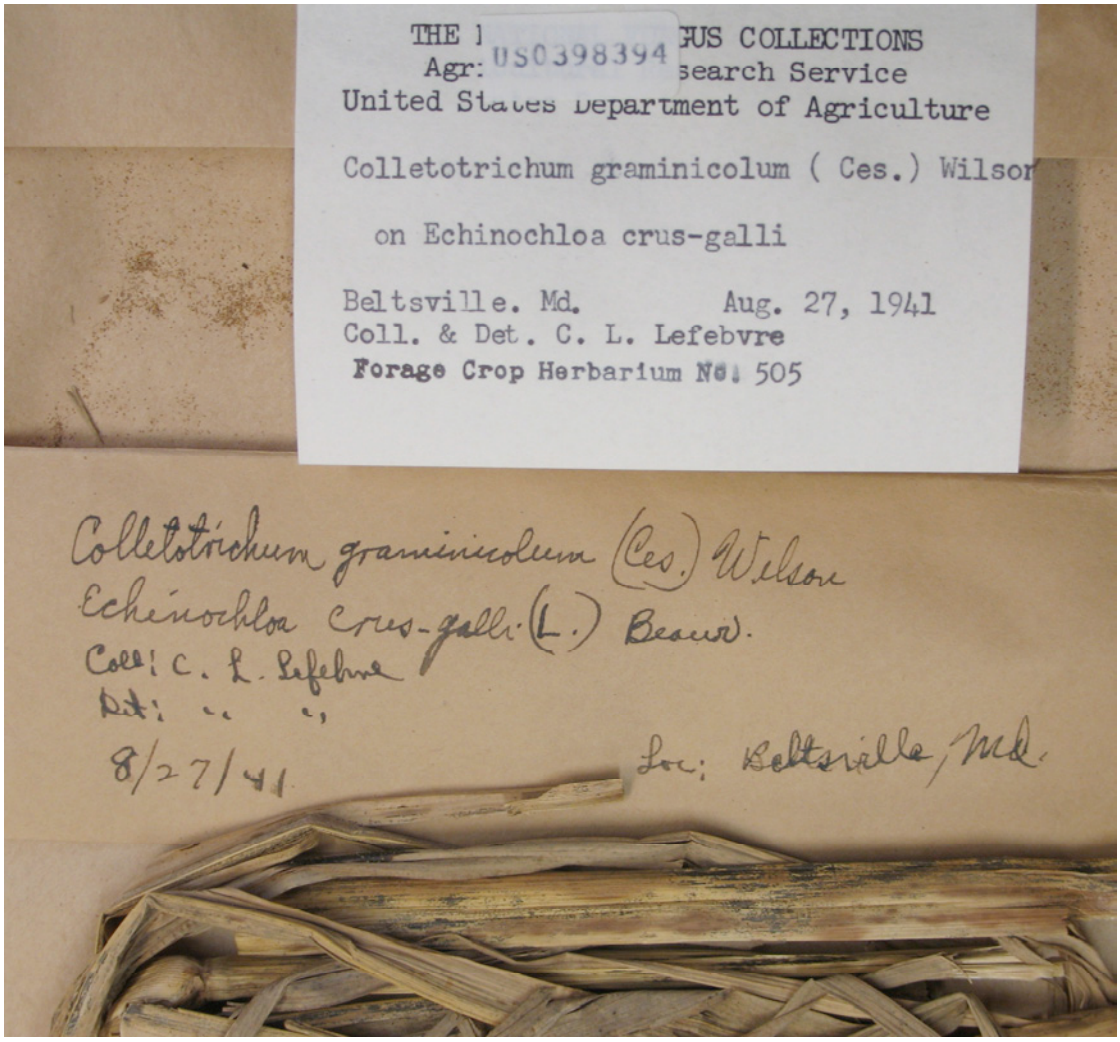


Figure 6.11.2

Herbarium specimens of *C. jacksonii*. (A) and (B) disease symptoms on infected *Echinochloa crus-galli*, (C) *C. jacksonii* setae, 400x magification. BPI398395, collected 25-Sept-1943.

(A)



(B)

Figure 6.11.3
Herbarium specimens of *C. jacksonii*. (A) and (B) disease symptoms on infected *Echinochloa crus-galli*. BPI398396, collected 27-Aug-1941.

(A)

(B)

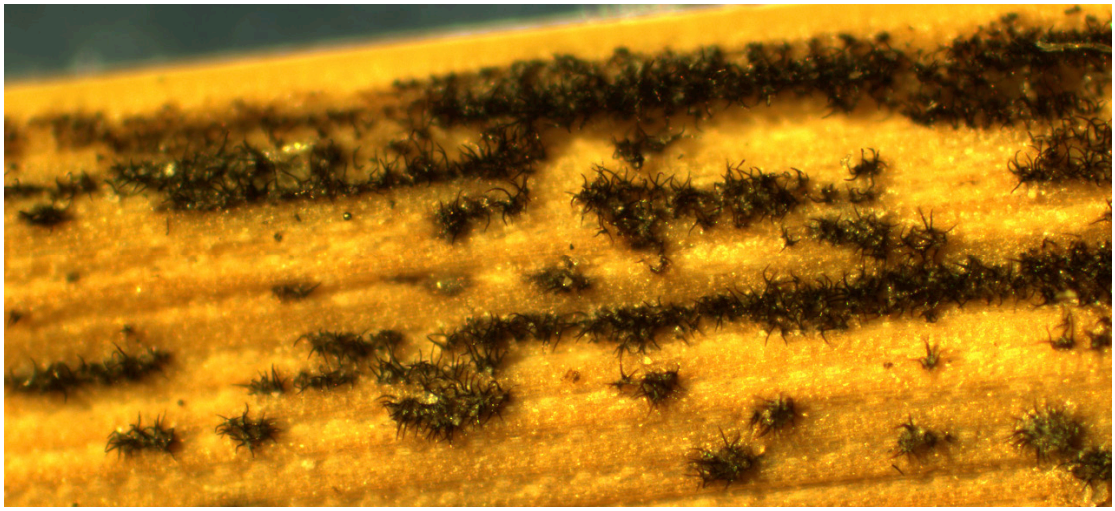


Figure 6.11.4
Herbarium specimens of *C. jacksonii*. (A) disease symptoms on infected *Echinochloa pungens* (B) *C. jacksonii* setae, 400x magification. . BPI398406, collected 26-Aug-1945.

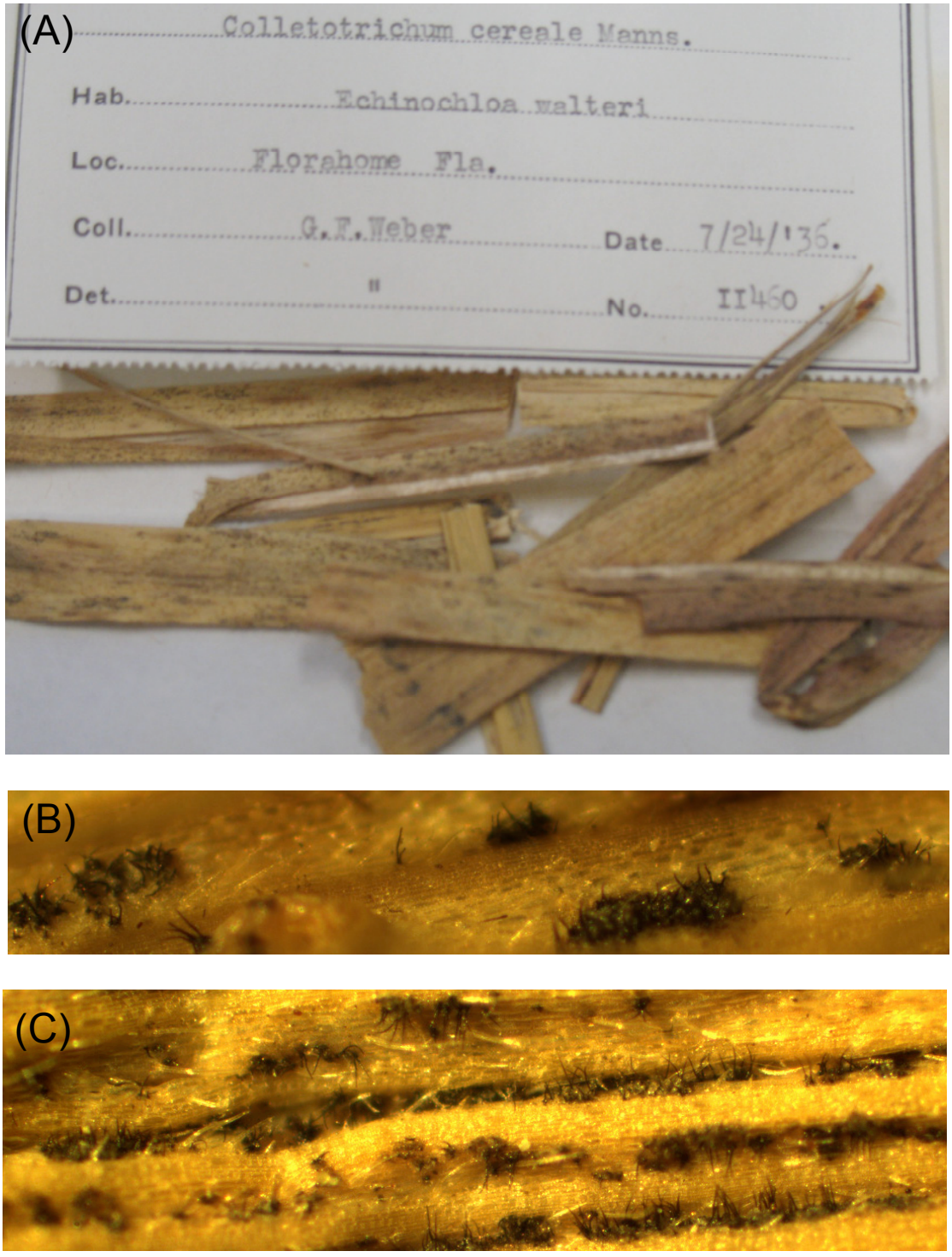
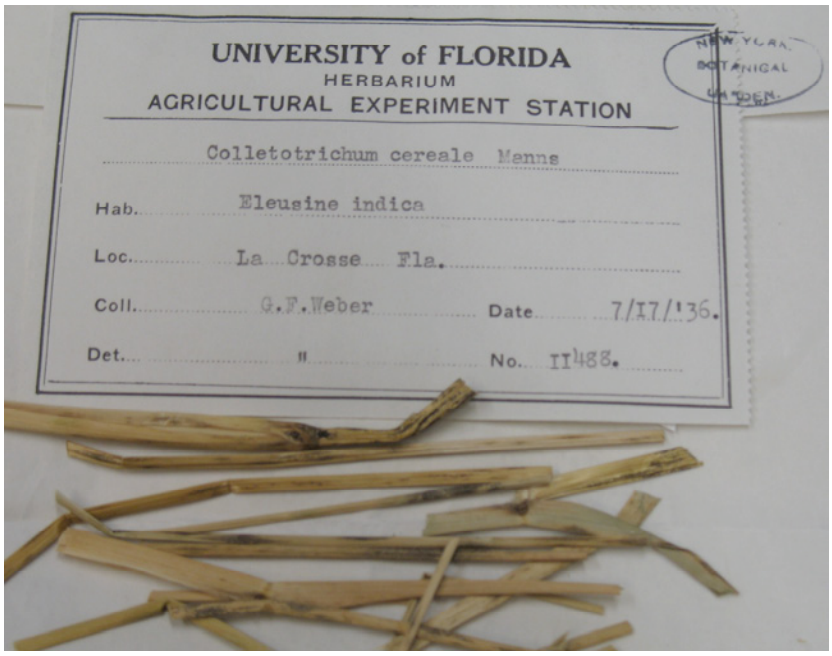
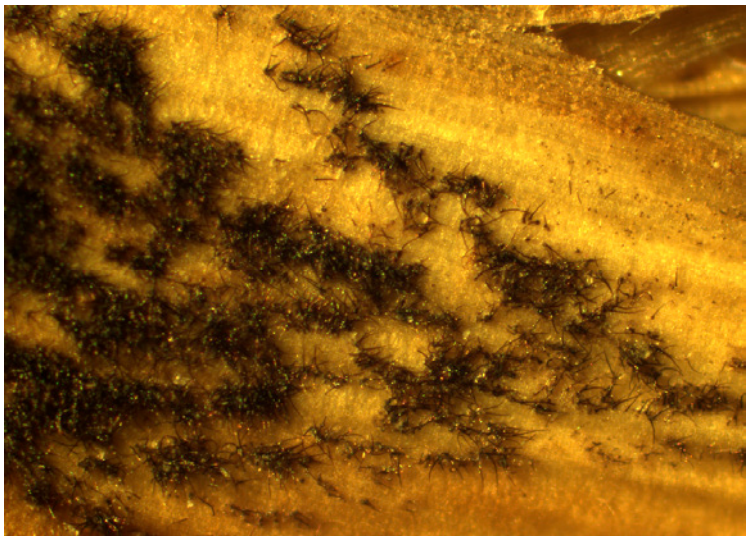


Figure 6.11.5
Herbarium specimens of *C. jacksonii*. (A) disease symptoms on infected *Echinochloa walteri* and (C) and (D) *C. jacksonii* setae, 400x magification. NYBG, collected 24-Jul-1936.

(A)



(B)



(C)

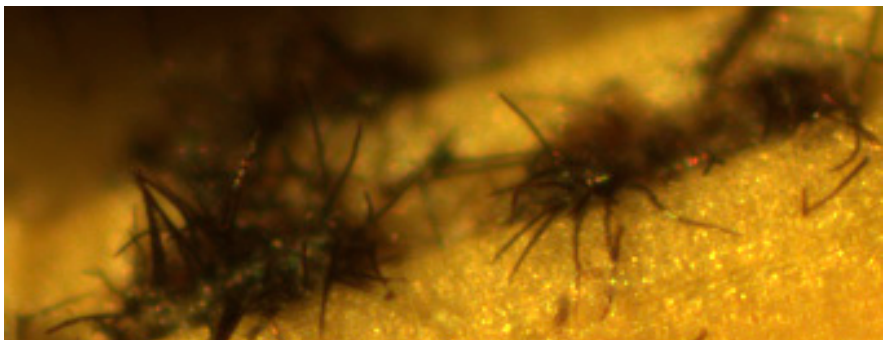


Figure 6.11.6
Herbarium specimens of *C. eleusines*. (A) disease symptoms on infected *Eleusine indica* and (B) (C) *C. eleusines* setae; NYBG, collected 07-Jul-1936.

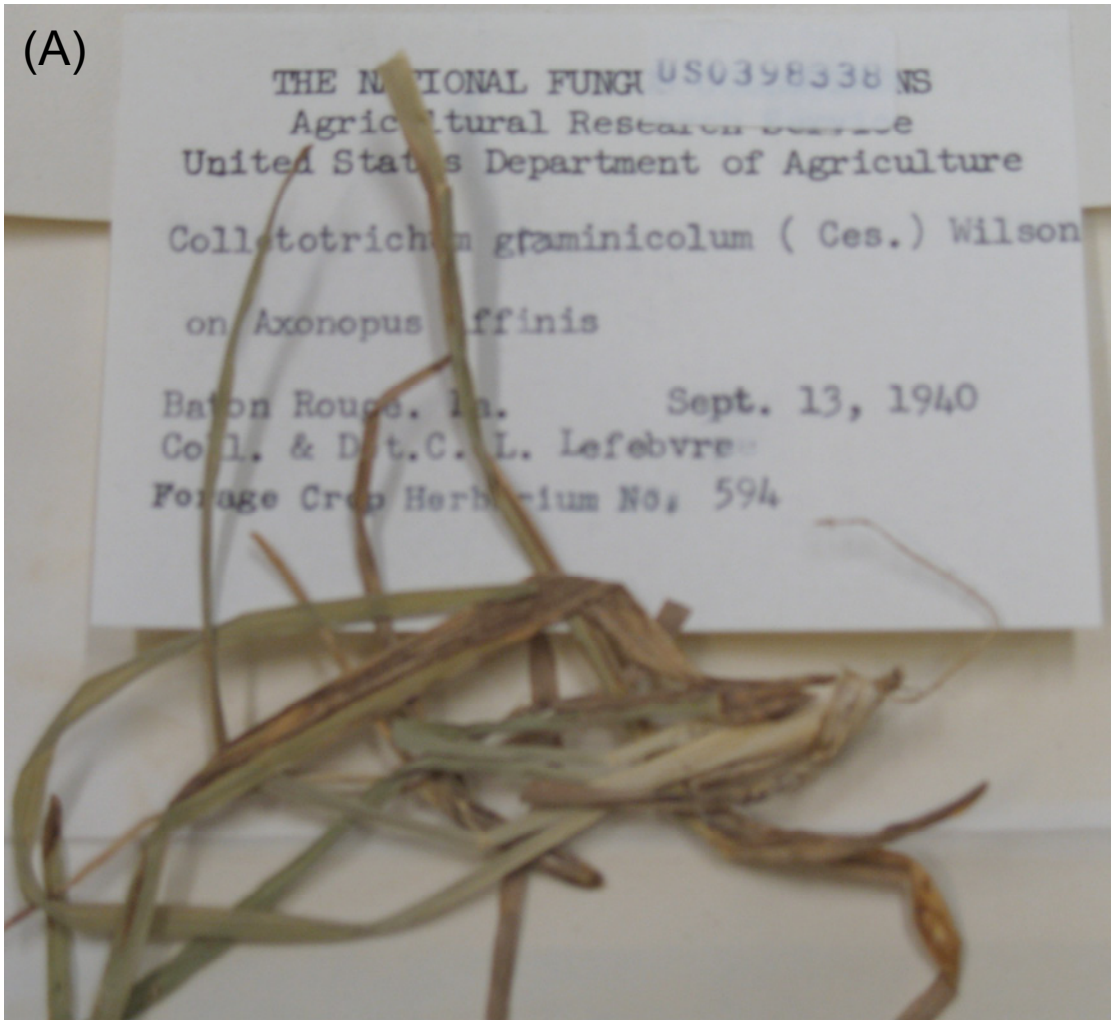


Figure 6.11.7
Herbarium specimens of *C. axonopodi*. (A)and (B) disease symptoms on infected *Axonopus affinis* BPI0398338, collected 13-Sept-1940.

(A)

(B)

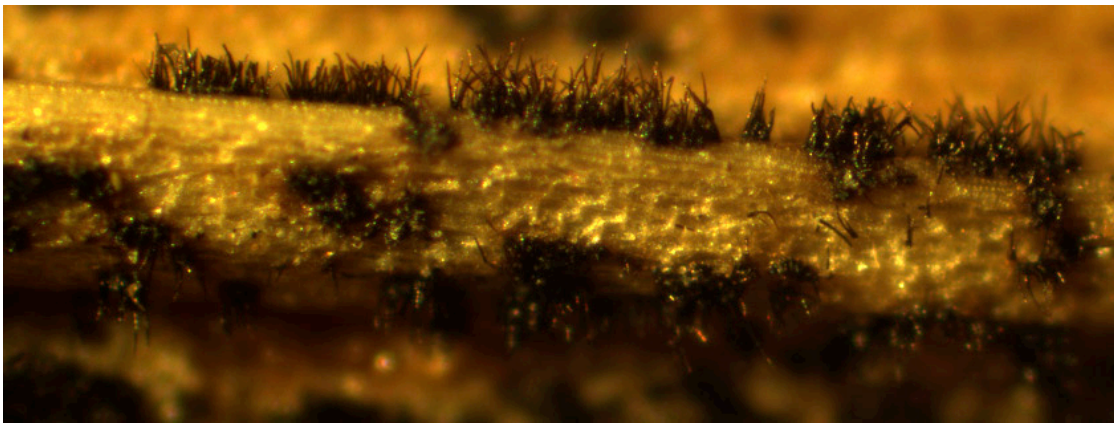
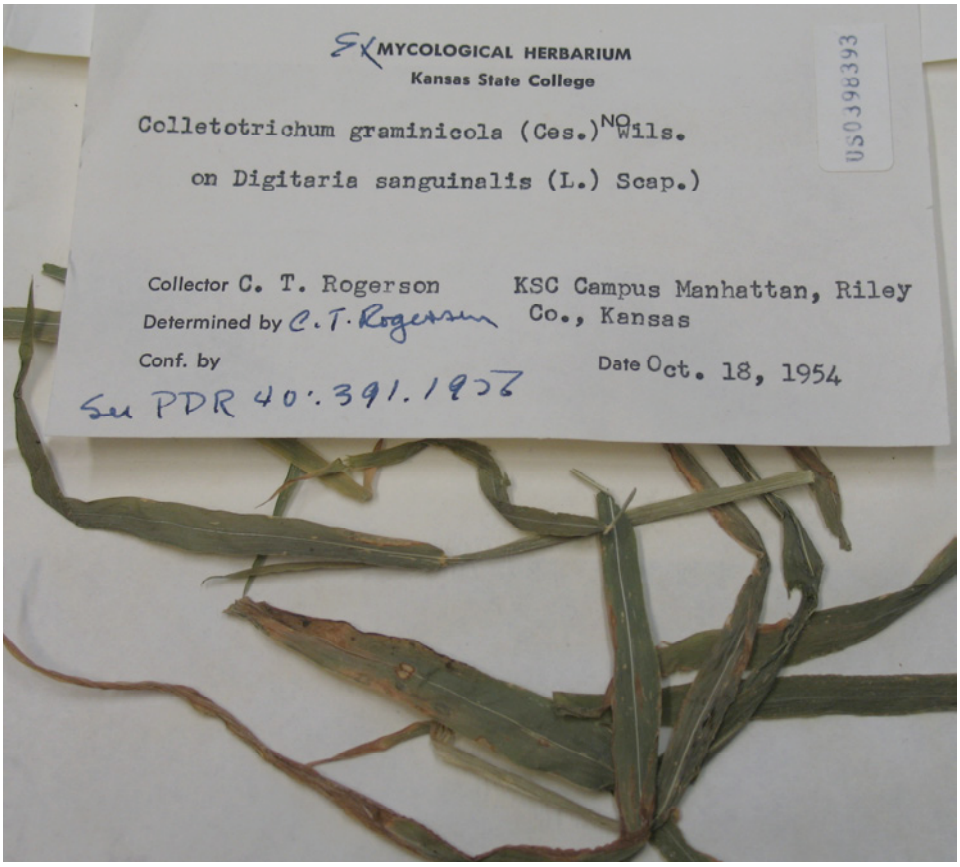


Figure 6.11.8
Herbarium specimens of *C. axonopodi*. (A) disease symptoms on infected *Axonopus compressus* and (D) *C. axonopodi* setae, 400x magification. BPI0398339, collected 10-Dec-1946.

(A)



(B)

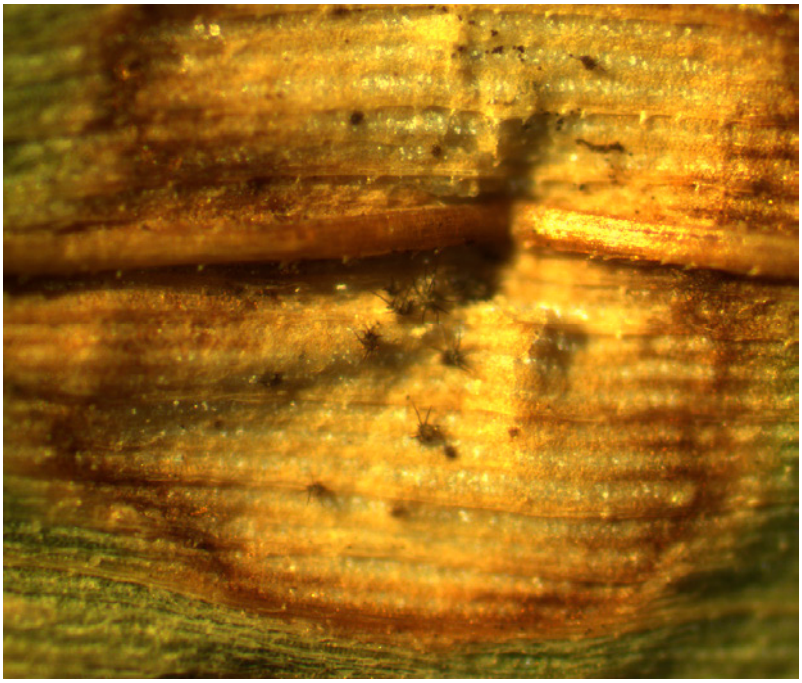
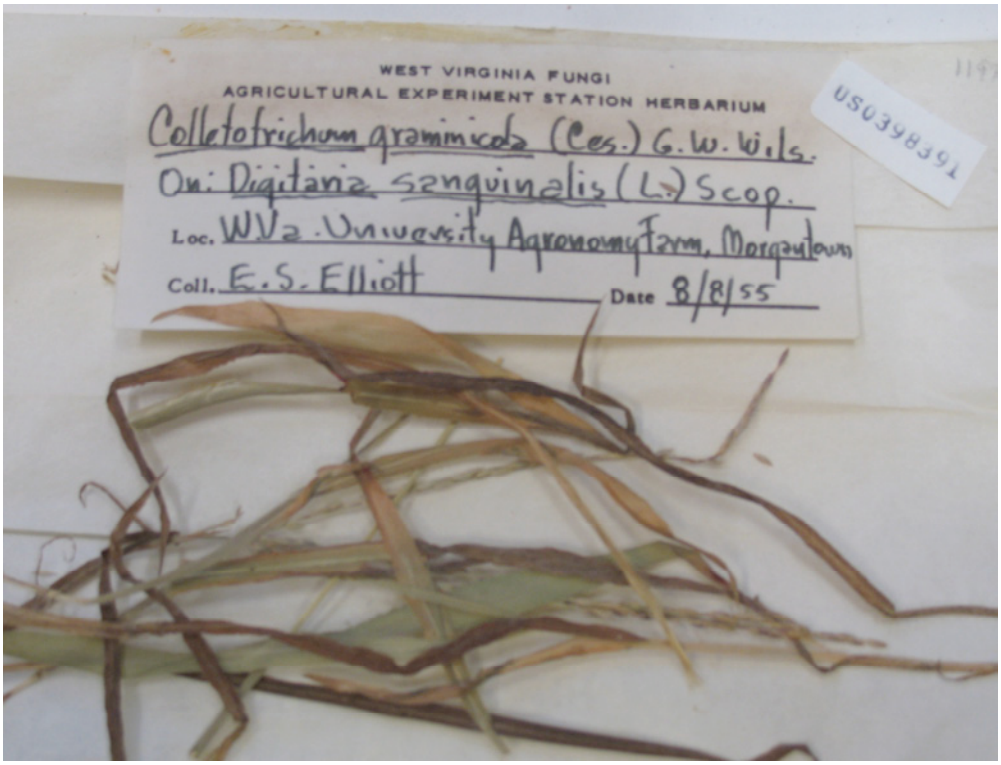


Figure 6.11.9
Herbarium specimens of *C. hanaui*. (A) disease symptoms on infected *Digitaria sanguinalis* and (D) *C. hanaui* setae, 400x magification. BPI0398393, collected 18-Oct-1954.

(A)



(B)

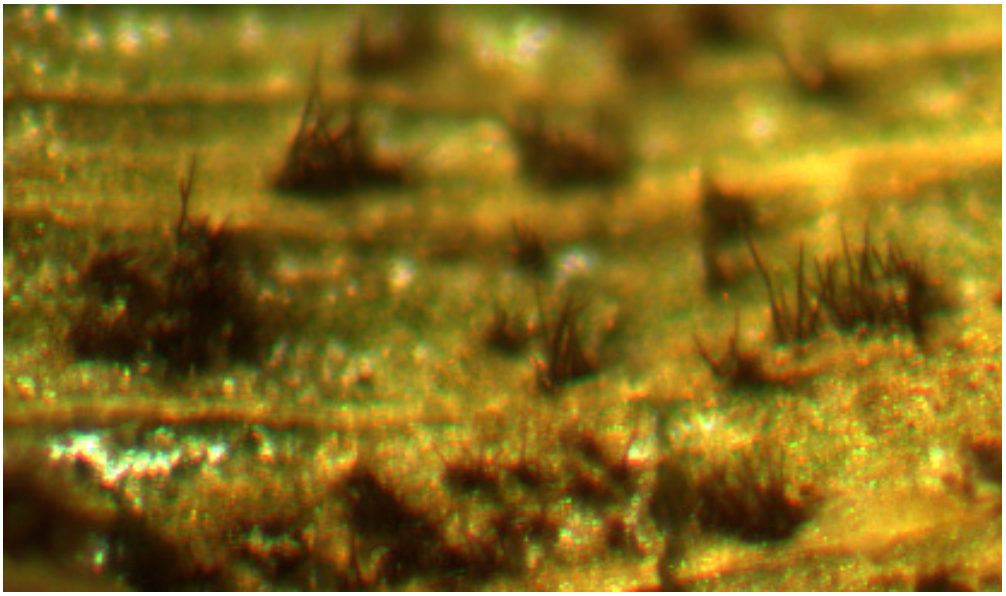


Figure 6.11.10
Herbarium specimens of *C. hanaui*. (A) disease symptoms on infected *Digitaria sanguinalis* and (D) *C. hanaui* setae, 400x magification. BPI0398391, collected 08-Aug-1955.

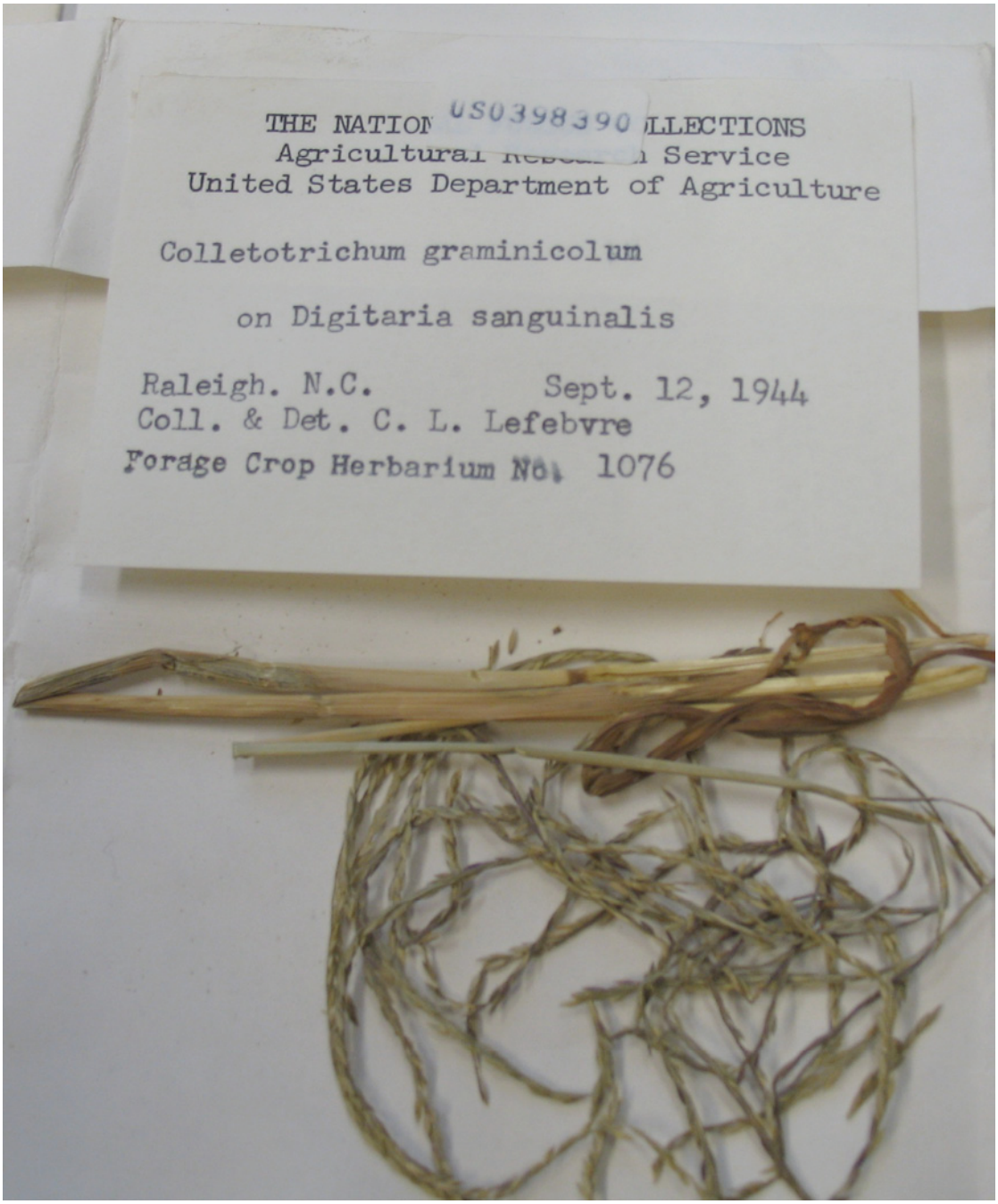


Figure 6.11.11
Herbarium specimens of *C. hanaui*. Disease symptoms on infected *Digitaria sanguinalis*, collected 12-Sept-1944.

Chapter 7: Can species of *Colletotrichum* be uniquely identified through ITS sequence data? A case study using the falcate-spored graminicolous *Colletotrichum* group

7.1 Introduction

Just over 50 years ago, in a landmark treatise, Josef von Arx ushered in the modern era of *Colletotrichum* systematics by decreasing the number of accepted species from over 750 to just 11 (von Arx, 1957), signaling an end to what was considered the outdated practice of describing a new species taxon to correspond with every novel host plant association (Fig 7.1). As early as the 1960s; however, challenges to the von Arx classification were advanced by Brian Sutton, a new student of mycology completing his doctoral research on *Colletotrichum* taxonomy at the University of London (Sutton, 1965). Throughout his career, Sutton maintained that although the von Arx *Colletotrichum* classification was an invaluable taxonomical work, founded as it was on the modern approach of morphologically-based species recognition, it was nevertheless flawed due to sweeping overgeneralizations (Sutton, 1965; 1966; 1968; 1980; 1992). The generality of the work likely resulted from the fact that von Arx examined only ~30% of the specimens he discussed, basing most of his conclusions on descriptions from the literature (von Arx, 1957). Sutton's critique has been repeatedly confirmed as valid in the intervening years. Several of the von Arxian synonomies, including *C. gloeosporioides*/*C. acutatum*, *C. graminicola*, *C. dematium*, and *C. capsici*, are exceptionally broad interpretations, and considered by Sutton and others to be species "groups" rather than distinct taxonomic entities (Sutton, 1980 and 1992; Cannon *et al*, 2000). In particular, both *C. gloeosporiodes*

and *C. dematium*, with >800 and 84 synonyms listed by von Arx, respectively, possessed such an extreme range of morphological and host variability that standardized descriptions and morphological keys would be meaningless.

Despite the 43 years that have passed since Sutton first pointed out the unresolved issues for the genus, *Colletotrichum* taxonomy is still dominated by the classification of von Arx. One of the primary reasons for the lack of progress in the field is the limited number of morphological characters available for taxonomic purposes. Many *Colletotrichum* species and strains do not produce the teleomorphic state, *Glomerella*, eliminating a wealth of informative criteria that might otherwise be evaluated from sexual morphological structures and tests of interfertility. The anamorphic *Colletotrichum* state produces few morphological structures, with only conidial and appresorium size and shape possessing any utility for interspecific taxonomic study. And even these morphological characters are limited in their resolving power, particularly the conidial features (Cannon *et al.*, 2000; Sutton, 1965,1980, 1992). The simplest distinction between intrageneric taxa is made through condia shape, with two major variants – straight or falcate. Outside of these two primary morphs, with few exceptions (i.e., *C. caudatum*) the size and shape of conidia are present in broadly overlapping ranges between species, and often exhibit a range of minor variability even on an intraspecific or individual level (i.e., *C. cereale*; Crouch *et al.*, 2006) that precludes conidal morphology as a taxonomic character except for deriving only the most general species comparisons. Hyphal appressoria (hyphopodia), although not universally produced by *Colletotrichum*, have been used to discriminate between a few species, most notably between *C. graminicola sensu stricto* Sutton and *C. sublineolum* (Sutton 1966, 1968). The precision of appressoria for taxonomic research is not altogether satisfactory, as recent comparisons of appressoria against species boundaries in the

systematically well-studied graminicolous *Colletotrichum* group showed that this character is homoplasic and inadequate for distinguishing between species except in rare cases (*i.e.*, *C. graminicola*) (Chapter 6).

In recent years, molecular tools have been employed to identify and characterize *Colletotrichum* species, beginning with the application of marker-based methodology in the late 1980s, followed by analysis of the internal transcribed spacer (ITS) gene trees beginning in the 1990s (Sherriff *et al*, 1994,1995; Sreenivasaprasad *et al.,* 1996), and, more recently, molecular phylogenies from two or more loci (Crouch *et al.,* 2006, 2008; Du *et al*, 2005; Farr *et al.,* 2006). With few exceptions, the use of the rDNA internal transcribed spacer region (ITS) sequence dominates molecular analyses of *Colletotrichum* performed to date, but there are concerns as to whether this sequence provides the accuracy and resolution needed to evaluate intrageneric taxa systematically. Three molecular phylogenetic analyses encompassing the entire genus have been completed, with each of these works based on a single genomic region – the ITS (Moriwaki *et al.*, 2002; Sherriff *et al.*, 1995; Sreenivasaprasad *et al.,* 1996). The resultant gene genealogies from these studies were poorly supported by bootstrap values at several key nodes, and resolution of species groups was not achieved. In particular, from the parsimony tree of Moriwaki *et al.,* (2002), only four ingroup nodes out of 16 total nodes were supported by bootstrap values greater than 50. Similar outcomes have been reported when the ITS gene was used to evaluate smaller sub-groups within the genus (for example, Crouch *et al*, 2005; Du *et al*, 2005). The results of previous research suggest that the ITS sequence cannot be used independently to address adequately the species delimitation problems espoused by Sutton, nor to reconstruct the evolutionary history of the genus *Colletotrichum*. But could the ITS sequence be used as a marker to identify already well characterized phylogenetic

species that have been circumscribed through robust multilocus phylogenetic frameworks? Given the large archive of *Colletotrichum/Glomerella* ITS sequences already deposited with the International Nucletoide Sequence Databases (1,433 sequences as of 20-Mar-2008) (INSD; GenBank, EMBL and DDBJ) and the wide spread usage of this sequence for species diagnosis, there is a real need to address the accuracy and utility of ITS sequence data.

In this study, I set out to determine the extent to which phylogenetic species of *Colletotrichum* correspond to ITS sequence data and gene genealogies. I hypothesized that despite the questionable nature of this character when used independently in systematic investigations of evolution and taxonomy in the genus, the ITS sequence might be employed *ex post facto* by mycologists, plant pathologists, and ecologists to identify phylogenetic species that have already been characterized using the more robust methodology of multilocus phylogenetics. With the recent movement within the mycological community to build a universal ITS-based barcoding platform for the fungal kingdom, addressing the reliability of the *Colletotrichum* ITS sequence is topical: can ITS sequences be developed into accurate, cost effective and easily implemented identification tools for the genus *Colletotrichum*?

To consider these questions, I test the correspondence of the ITS sequence against phylogenetic species diagnosed using an implementation of the genealogical concordance phylogenetic species method described by Taylor and colleagues (Dettman *et al.*, 2003; Taylor *et al.*, 2000). For this research, the falcate-spored, graminicolous (FG) group of *Colletotrichum* species is used as a model for the genus, as it possesses several advantages that make it an ideal system to investigate the questions proposed as part of this work. First, unlike the other major *Colletotrichum* group species, the FG group (the so-called

"*graminicola*" group of von Arx) has been extensively analyzed, with species and population-level distinctions clearly defined through multilocus phylogenetic and population genetic analyses (Crouch *et al.,* 2006; 2008a, 2008b; chapters 1, 4 and 6). In the von Arxian treatment of the FG group, all graminicolous isolates of *Colletotrichum* were all included within the circumscription of *C. graminicola* G.W. Wilson (von Arx, 1957). Sutton separated *C. falcatum* and *C. sublineolum* from *C. gramincola sensu lato* von Arx using appressorial morphology and *C. caudatum* was distinguished through the presence of a unique conidial appendage (Sutton 1965, 1966, 1980, 1992). Using multilocus phylogenetics, Crouch *et al.,* further separated *C. cereale* (2006) and *C. eleusines* (chapter 6) from *C. graminicola sensu stricto* Sutton and erected six additional *Colletotrichum* species from grass hosts (chapter 6). Thus, the FG *Colletotrichum* are systematically well defined, and at present, are the only major von Arxian "group species" where species boundaries have been confidently resolved. Secondly, with few exceptions, species of the FG group maintain a close alliance with their hosts, with individual species limited to a single grass genus or species (Crouch *et al.,* 2006 and 2008; chapters 1, 4, 6). Even the generalist FG species *C. cereale*, with a host range of at least 27 grasses, is subdivided into populations that are largely defined by lifestyle and/or host plant (Crouch *et al*., 2008; chapter 4). Importantly, it was observed that in the FG group, both species and populations could be defined with precision using the same four gene, multilocus dataset (Crouch *et al*., 2008; chapters 4 and 6). With this fundamental understanding of the *Colletotrichum* FG group underpinning the investigation, in this paper I analyze the efficacy of ITS sequence data in distinguishing *Colletotrichum* taxa.

7.2 Materials and Methods

7.2.1 Fungal isolates

A total of 55 specimens of falcate-spored *Colletrichum* isolated from grass hosts (the FG group) were assessed for this study (Table 7.1). In addition, 17 samples of *Colletorichum* sharing a common falcate-shared spore morphology (the FN group, refering to falcate-spored, non-graminicolous) with the FG *Colletotrichum* were included in the ITS analysis (Table 7.2). Although the FN samples of non-graminicolous *Colletotrichum* isolates in this study presumably represent four falcate-spored species described using morphological characters (*C. dematium*, *C. trichellum*, *C. truncatum*, *C. trifolii*), in many cases their taxonomy is uncertain (Sutton, 1982 and 1990), and species assignments for the purposes of this work are based upon names under which these isolates were curated in the culture collections and evaluation of cultural and morphological characteristics (chapter 6), not systematic analyses. Fungi were grown and maintained as previously described (Crouch *et al.*, 2006).

7.2.2 Molecular phylogenetics

ITS sequence data were generated for 17 FN taxa (Table 7.2) as described by White *et al.*, (1999). *C. acutaum*, a species of *Colletotrichum* with oval spores, was included as the outgroup taxon (Crouch *et al.*, 2006; chapter 1). Alignment of the new FN sequences with existing FG sequences (chapters 4 and 6) was performed using Clustal W (Thompson, 1994) on the EBI/EMBL site (http://www.ebi.ac.uk/Tools/clustalw), then

adjusted to remove all gap sequences and ambiguously aligned regions. The resultant 398-bp dataset was analyzed using maximum likelihood (ML) in PAUP* (Swofford, 2000) modeled on the parameters estimated using Modeltest v.3.06 (Posada and Crandall, 1998), with the analysis performed as previously described (chapter 6). Bayesian posterior probabilities in support of the ML topologies were estimated in MrBayes v.3.0b4 (Huelsenbeck and Ronquist, 2001), as described (chapter 6).

7.2.3 Evaluation of ITS sequences curated by the International Nucleotide Sequence Database (INSD; including GenBank, EMBL and DDBJ)

Representation and characteristics of ITS sequences derived from isolates of *Colletotrichum* and *Glomerella* curated by the INSD were evaluated by searching the core nucleotide sequences through NCBI GenBank for these genera using "internal transcribed spacer" (search of database performed 20-Mar-2008). ITS sequences generated and accessioned on GenBank through the present study were not included as part of the search. ITS sequence accessions were counted as either "published" or "unpublished", referring to whether or not the sequence was associated with a peer-reviewed publication and/or a voucher specimen.

All ITS sequences generated in the present study were used to search the INSD through GenBank using nucleotide mega-Blast searches (Altschul, 1990), with a preliminary species identification performed based on the best matches from GenBank accessions in order to estimate the ability of the ITS sequence to accurately pinpoint

species identity using database matches, e-values, similarity scores, and through the distance tree of search results constructed by the GenBank interface (minimum evolution option).

7.3 Results

7.3.1 The ITS gene genealogy as a phylogenetic indicator of species boundaries and supraspecific groups

The ITS gene genealogy for the FG *Colletotrichum* (Fig. 7.2), when compared against the four gene, three locus phylogeny published for the group (Fig. 6.1; *Sod2, Mat1-1*, *Apn1* and ITS), showed an appreciable loss of resolution across both species boundaries and supraspecific groupings. The multi-gene tree topology was well-supported at four levels and minutely reflected the history of this group, ranging from the most basal, supraspecific nodes where the FG group diverged between the C3 and C4 lineages, to the supraspecific clades that unite many of the sister C4 taxa, to the species boundaries that largely correspond to host origin, and even to the tip nodes corresponding to subspecific populations. In contrast, the ITS tree did not even recover the major subdivision of the group into the C3 and C4 lineages (Fig 7.2; chapters 4 and 6). Supraspecific structuring of the C4 species (*C. graminicola/C. paspalum*; *C. falcatum/C. caudatum/C. miscanthi/C. sublineolum/C. eleusines*, chapters 4 and 6) was also lost in the ITS tree, with only the *C. hanauii/C. jacksonii* group identified. In addition, species identification was inconsistent across the ITS tree: *C. graminicola*, *C. sublineolum*, *C. jacksonii*, and *C. falcatum* were all

represented and supported by posterior probabilities, but the remaining FG species – most notably *C. cereale* and the *C. nicholsonii* group – did not exist, with or without support. Population subdivisions identified through the multi-locus phylogeny and other gene trees derived from the *Mat1-1*, *Apn1* and *Sod2* (Crouch *et al*, 2008; chapter 4) were not present in the ITS tree.

An unanticipated result from the ITS tree was the association of two *C. cereale* isolates (KS-TA4-F4 and NJ-DG-2A2) within a well supported group of the FN *C. dematium* isolates; this grouping may play some role in the lack of cohesion amongst the *C. cereale* sequences. In a previous analysis of the ITS region where only FG isolates were considered using identical phylogenetic methodology, *C. cereale* was identified as a unified taxon, albeit with very low bootstrap support (<50; *e.g.* chapter 4). Blast searches performed using these two *C. cereale* sequences showed them to be most closely allied with INSD sequences derived from either *C. dematium* or *C. truncatum* rather than *C. cereale* (Table 7.3), raising questions about the evolutionary relationship between the FN and FG group. The FN group, comprised of *C. dematium* and eight other species of falcate-spored *Colletorichum* associated with non-grass host plants, is one of the most poorly understood taxonomic groups in the genus. Although the FG and FN groups share a common conidial morphology, it is currently unknown whether that trait is evolutionarily conserved, or has independent origins. The ITS data from this study suggest the possibility that the two falcate spored groups may be evolutionarily related to one another more closely than expected, but is insufficient for the purpose of testing or expanding upon the idea.

A second well-supported group was inferred from the FN taxa, distinct from the FG group. This group was comprised of four isolates of *C. dematium* and one *C. truncatum*;

however, Blast searches of GenBank identified these sequences as *C. capsici*. This similarity-based identification is inconsistent with spore morphology, as all five of these isolates possess the narrow, strongly tapered falcate spores characteristic of the *C. dematium* group, rather than the wider, less acutely apiced spores of *C. capsici* (Fig. 6.3; Sutton, 1980).

The absence of basal resolution in the ITS tree is in accordance with the findings of previous studies in the genus where species level and even population boundaries could be resolved with this sequence, while older, supraspecific relationships remained uncertain. This likely reflects the high variability of the ITS, and the loss of numerous nucleotide characters due to regions of ambiguous alignment, or saturation of the sequences. For example, from sequences that were originally between 600-bp and 550-bp long, it was possible to include with confidence only 398-bp in the alignment in the present study, as the inclusion of the outgroup taxa and the FN species introduced numerous uncertainties about positional homology which could not be satisfactorily resolved. But in a previous study where the same dataset was analyzed without the 17 FN taxa, it was possible to retain 450-bp for analysis (chapter 4). In earlier work, where only *C. cereale*, *C. graminicola* and *C. sublineolum* were sampled, 482-bp were preserved (Crouch *et al*, 2006; chapter 1). Consequently, between12-18% of the characters were eliminated from the present analysis relative to previous studies of the FG group through the inclusion of the FN taxa. These data demonstrate that when divergent species are included in the same dataset, some of the most variable and informative characters in the resultant ITS alignment will by necessity be sacrificed for the sake of accuracy.

7.3.2 Database searches using ITS sequences from authentic FG

Colletotrichum species

With only a few exceptions, the majority of credible matches to the Blast searches were made to sequences derived from falcate-spored taxa, either of the FG or FN groups (Table 7.3). Ranking the "best" matches using standard measures of similarity such as e-values and maximum identity had limited application in determing the actual species from which the ITS sequences were derived. All e-values for the 100 Blast hits generated by the searches were significant (0.00) and maximum sequence identity values ranged from 90-100%. Maximum sequence identity values from matches with non-target species were typically within zero to four percentage points of sequences from the actual *Colletotrichum* species. Because some species are overrepresented or underrepresented on the database (Table 7.4), the total number of hits per species was not taken into account when determining species identity. For these reasons, final species identification was made from the distance tree of the results. If the query sequence was not placed within a group in the distance tree, or if the group was comprised of multiple species, then maximum identity was used for clarification and final determination. As the objective of this research is to test the ITS region as a tool for identification purposes, only sequence data were considered for this analysis, and host range criteria were not taken into account when delimiting species.

Blast searches performed using the *Colletotrichum* FG ITS sequences led to mixed results for accurate species diagnosis (Table 7.3). Sequences from authentic isolates of *C. graminicola*, *C. sublineolum* and *C. caudatum* best corresponded to the ITS sequences of their respective species, with maximum sequence identities between 90-92%. *C. falcatum* isolates were split between *C. graminicola* and *C. falcatum* database sequences, perhaps

reflecting the diverse lineages of this polyphyletic species (chapter 6). ITS sequences from the six novel FG taxa that were recently described but have not yet been accessioned on GenBank (*C. hanaui*, *C. nicholsonii*, *C. jacksonii*, *C. axonopodi*, *C. eleusines* and *C. miscanthi*; chapter 6) were not expected to yield good matches, but several reasonable species diagnosis were made from the ITS sequences. Without previous knowledge of their identity, the matches would have been accepted as legitimate. *C. nicholsonii*, *C. hanaui* and *C. eleusines* were well matched with *C. gramincola* sequences, while *C. jacksonii* and *C. miscanthi* sequences were consistent with sequences curated as either *C. falcatum* or *C. gramincola*, and *C. axononpodi* was identified as *C. sublineolum* (Table 7.3).

Eight of the eleven *C. cereale* ITS sequences were best matched against *C. gramiincola* sequences from the database (Table 7.4). The erroneous matches resulted either from matches with actual sequences of *C. graminicola*, or the presence of ITS sequences from *C. cereale* that were incorrectly named as *C. graminicola*. The *C. cereale* sequences misnamed as *C. graminicola* were present in the database due either to misdiagnosis, or pre-2006 sequences that were never updated to reflect the 2006 name change (Crouch *et al*, 2006). Allowing third party annotation of GenBank sequence, as lobbied for by the mycology community (Bruns *et al.*, 2008), would eliminate the errors in identification due to the outdated records. As described above (section 7.3.1) two *C. cereale* isolates were identified as either *C. dematium* or *C. truncatum*, and only a single strain (NJ-8626) was accurately identified as *C. cereale* using ITS sequence searches of the database.

In their analysis of all fungal ITS sequences on the INSD databases, Nilsson *et al.*, (2006) determined that more than 10% of the fungal ITS sequences have deficient taxonomic annotations, and lack the information required to make informed decisions as

to whether the name assigned to the sequence is accurate or not. I searched the INSD databases to establish a general benchmark for the accuracy for *Colletotrichum* ITS sequences and found them to be overrepresented in the public databases relative to all other genes, with 1444 of the 3185 total sequences curated by the INSD drawn from this one region (45%; Table 7.4; search performed 20-Mar-2008). The greatest concentration of ITS sequences emanated from just two species: *C. gloeosporioides* (*G. cinulata*) and *C. acutatum* (*G. acutata*), which comprised 54% of all ITS sequences curated on the database (Table 7.4). 39% of the 1444 ITS sequences existed only as sequence data (Table 7.4) within the database, and did not correspond with authentic, characterized specimens or publications in the literature (Table 7.4). The *Colletorichum* data were consistent with findings reported in a recent publication, where 40% of all fungal kingdom ITS sequences deposited in the database consortium were not associated with publications (Nilsson *et al*, 2006). For ITS sequences cataloged under the teleomorphic *Glomerella* name, the proportion of unpublished sequences was substantially larger than that reported for all fungi, with 50% of the 844 *Glomerella* sequences curated exclusive of publication and/or voucher specimens. The percentage of unpublished *Colletotrichum/Glomerella* ITS sequences curated by the INSD databases (47%) exceeded even the alarmingly large numbers reported for the fungal kingdom as a whole.

7.4 Discussion and Conclusions

In one important way this study confirms something already well known about the ITS sequence that this region should not be used independently to infer evolutionary

relationships in the genus *Colletotrichum*, nor should it be employed to describe novel *Colletotrichum* taxa, regardless of rank. It is well established that the ITS sequence often generates poorly supported, unresolved phylogenetic trees, particularly when distant taxa are sampled. There are also numerous examples in the literature where single gene trees are shown as incongruent with the actual species tree. In the present work, for example, six individuals drawn from four species (Fig. 7.2) were positioned outside the context of their known taxonomic/evolutionary environments in the ITS gene tree (Fig. 6.1; chapter 4 and 6).

During the 15 years that this region has dominated *Colletotrichum* systematics, the von Arxian "group species" dilemma has not been solved, and the questions put forward by Sutton in 1965 remain largely unanswered. But Sutton's arguments were correct, and the von Arxian taxonomy is overly general, as exemplified in the recent resolution of the "*graminicola*" group, where multigene, multilocus molecular phylogenetics uncovered 14 species (chapter 6) where there was once only a single entity. The failure of the ITS to address *Colletotrichum* taxonomy sufficiently, like the failure of morphological and host range criteria before it, did not come about because no solution to the von Arxian "group species" problem exists. Rather, there is simply not enough signal in the ITS region for the task of systematically resolving the taxonomy and evolution of the genus *Colletotrichum*.

Because *Colletotrichum* is an important genus of plant pathogens that cause disease in an estimated 900 plant species worldwide, the question of species identification goes beyond academic pursuits such as reconstructing evolutionary history, understanding population processes and taxonomic inquiry. For *Colletotrichum*, species identitification is tied to practical considerations such as effective disease control, quarantine measures, economic losses, and in extreme cases, adequate food supply. Despite persistent questions

about the reliability of morphological, host range and ITS characters to provide accurate species-level identifications in the genus *Colletotrichum*, these criteria are currently the only substantive tools available to the plant pathology community for diagnostic purposes. Between 1997-2008, 53 peer-reviewed reports of *Colletotrichum* causing disease in novel plant hosts were published in the journal *Plant Disease*, the primary avenue for reporting novel host/pathogen associations, geographic locations, and/or plant diseases. As shown in Table 7.5, of the 53 *Colletotrichum* disease reports, 72% relied entirely upon morphological criteria such as colony characteristics and spore shape/size, while 28% used a combination of morphology and the similarity of ITS sequences to data curated on GenBank. These criteria make up the bulk of the toolkit presently available to plant pathologists and health care practioners. Recent work in the *Colletorichum* FG group demonstrated that morphological characters are imprecise for species identification purposes (chapter 6). As established here, the use of ITS sequence data for species diagnosis may also encounter difficulties, even when well-studied, authentic specimens are evaluated. But the findings of the present study, where the integrity of the ITS sequence was tested within the context of a taxonomically well-characterized group of *Colletotrichum*, suggests that the ITS sequence tool, although in urgent need of repair, should not be entirely discarded.

How accurate are ITS-based similarity searches of public databases like GenBank for the diagnosis of *Colletotrichum* species? These results show that Blast searches of INSD sequence data are currently an unreliable method of identifying *Colletotrichum* species, even in a well sampled and extensively studied group such as the FG. At best, accurate species identities were generated through the Blast searches, and this occurred often for some of best represented and characterized species such as *C. graminicola* and *C. sublineolum*. At

worst, recognized species were misidentified even in the presence of conspecifics, and the presence of novel species went undetected. In these instances, representation in the databases and systematic characterization was no guarantee that accurate identifications would be made through database searches, as shown with the *C. cereale* Blast results. *C. cereale* was rarely matched to sequence accessions identified as *C. cereale*, even though this organism has 91 authenticated ITS sequences curated on GenBank, making it the third most represented *Colletorichum/Glomerella* species for ITS data (Table 7.4).

For the FG group, imprecise species identifications arising from the use of the ITS sequence are clearly not insurmountable, as host range criteria can be used to supplement diagnoses (chapter 6). The results of this research leads to the conclusion that, at least for the FG *Colletotrichum*, ITS-based identification is no better than host association criteria (chapter 6), and in almost all instances, ITS sequence analysis will not perform as well as a knowledgeable assessment of the afflicted host plant against the published literature. But for other groups of *Colletotrichum* such a distinction will be impossible, as many plants serve as host to two or more species of *Colletotrichum*, sometimes simultaneously. Therefore, we are back to our original question: can the ITS be used independently for the purpose of species identification and/or barcoding type applications?

For understudied and taxonomically uncertain groups such as the FN, the prospects of accurate "species" identifications using ITS sequences is fraught with problems that go beyond sequence similarity issues. For example, in the present study, five *Colletotrichum* isolates morphologically consistent with *C. dematium* were unambiguously identified as *C. capsici* through ITS searches of the public database (Table 7.3). But was the identification really unambiguous? The *C. capsici* ITS sequences on GenBank that supported the *C. capsici* diagnosis are exclusive to GenBank, and were submitted by five

different research groups without accompanying publication or voucher specimen (sequences include EU315004, DQ195696, EU315004, EF556207, DQ415651) over a period between Oct 2005 and Jan 2008. As documented by Nilsson *et al.*, (2006) the presence of such insufficiently identified sequences in public databases perpetuates a chain of mistaken identifications and erroneous classifications. Did the original *"C. capsici"* sequence accessioned in Oct 2005 (EU315004) lead to the propagation of *C. capsici* identifications for the more recent sequence deposits? How was the original *C. capsici* strain of Oct 2005 identified? The subsequent strains? Do any of these diagnoses reflect true *C. capsici*? Without associated publication details, it is impossible to determine the legitimacy of the identification at any level. Naming the five *Colletotrichum* isolates from this research as *C. capsisci* based upon these search results would be irresponsibe, but naming and accessioning the isolates as *C. dematium* only through condial morphology introduces a sequence dataset that is insufficiently identified, and that is equally unsatisfactory.

What is the solution? There are many benefits from the use of the ITS region for barcoding and other identification purposes that terminating use of the sequence is not recommended. Continued use of the ITS region is vital to ensure connectivity between *Colletotrichum* and the rest of the fungal kingdom, as mycologists are working to implement the sequence as a standardized barcoding tool. This will be particularly true for metagenomic and other ecological studies, as *Colletotrichum* species are ubiquitous in plant communities as both endophytes and pathogens. I offer several recommendations, that might serve to revive the *Colletotrichum* ITS sequence and transform it into a functional and versatile tool.

(1) Always include the ITS sequence in multi-gene, multi-locus systematic

analyses, to connect the ITS database with authentic *Colletotrichum* isolates that can serve as points of reference.

(2) Establish a curated, highly annotated *Colletotrichum* database of ITS and other sequence data generated through published phylogenetic and taxonomic studies, modeled on the Swiss-Prot or UNITE database.

(3) Establish a virtual culture collection to allow researchers to gain access to authentic cultures.

(4) Develop a second sequence to serve as an adjunct to the ITS sequence to expand the range of ITS for identification. The 3' end of the DNA lyase (*Apn1*), is a likely candidate gene for this purpose (Crouch *et al*, 2008; chapters 4 and 6).

7.5 References

Altschul, SF, Gish, W, Miller, W, Myers, EW, and Lipman, DJ. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.

Arie, T, Christiansen, SK, Yoder, OC, and Turgeon, BG. 1997. Efficient cloning of ascomycete mating genes by PCR amplification of the conserved *MAT* HMG box. Fungal Genet. Biol. 21:118-130.

Arx, JA v. 1957. Die arten der gattung *Colletotrichum* Cda. Phytopath. Z. 29:413-468.

Bruns, TD, Blackwell, M, Edwards, I, Taylor, AFS, Horton, T, Zhang, Kõljalg,N, May,G, Kuyper,TW, Bever, JD, Gilbert,G, Taylor, JW, DeSantis, TZ, Pringle, A, Borneman, J, Thorn, G, Berbee, M, Mueller,GM, Andersen, GL, Vellinga, EC, Branco, S, Anderson, I, Dickie, IA, Avis, P, Timonen,S, Kjøller, R, Lodge,DJ, Bateman,RM, Purvis, A, Crous, PW, Hawkes, CW, Barraclough,T, Burt, A, Nilsson,RH, Larsson,K-H, Alexander, I, Moncalvo, J-M, Berube, J, Spatafora, J, Lumbsch, HT, Blair, JE, Suh, S-O, Pfister,DH, Binder, M, Boehm, EW, Kohn,LM, Mata, JM, Dyer, P, Sung, G-H, Dentinger, B, Simmons, EG, Baird,RE, Volk, TJ, Perry, BA, Kerrigan, RW, Campbell, J, Rajesh, J, Reynolds, DR, Geiser, D, Humber, RA, Hausmann, N, Szaro, T, Stajich, J, Gathman, A, Peay, KG, Henkel,T, Robinson, CH, Pukkila,CJ, Nguyen, NH, Villalta, C, Kennedy,C, Bergemann, S, Aime, MC, Kauff, F, Porras-Alfaro, A, Gueidan, C, Beck, A, Andersen, B, Marek, S, Crouch, JA, Kerrigan, J, Ristaino, JB, Hodge, KT, Kuldau, G, Samuels,GJ, Raja, HA Hermann,Voglmayr, L, Gardes, M, Janos, DP, Rogers, JD, Cannon, P, Woolfolk, SW, Kistler,HC, Castellano, MA, Sandra L, Maldonado-Ramírez, M, Kirk, PM, Farrar, JJ, Osmundson, T, Currah, RS, Vujanovic, V, Chen, W, Korf, RP, Atallah, ZK, Harrison, KJ, Guarro, J, Bates,ST, Bonello, PE ,Bridge, P, Schell, W, Rossi, W, Stenlid, J, Frisvad, JC, Miller, RM, Baker, SE, Hallen, HE, Janso, J, Wilson, AW, Conway, KE, Egerton-Warburton, L, Wang, Z, Eastburn, D, Ho, WWH, Kroken, S, Stadler, M, Turgeon, G, Lichtwardt, RW, Stewart, EL, Wedin, M, Li, D-W, Uchida, JY, Jumpponen, A, Deckert, RJ, Beker, HJ, Rogers,SO, Xu, J, Johnston, P, Shoemaker, RA, Liu,M, Marques, G, Summerell, B, Sokolski, S, Thrane, U, Widden, P, Bruhn, JN, Bianchinotti, V, Tuthill, D, Baroni, TJ, Barron,G, Hosaka, K, Jewell, K, Piepenbring, M, Sullivan, RF, Griffith, GW, Bradley, SW, Aoki, T, Yoder, WT, Ju, Y-M, Berch, S, Trappe, M, Duan, W, Bonito, G, Taber, RA, Coelho, G, Bills, G, Ganley, A, Agerer, R, Nagy, L, Roy, BA, Læssøe, T, Hallenberg, N, Tichy, H-V, Stalpers, J, Langer, E, Scholler, M, Krueger, D, Pacioni, G, Pöder, R, Pennanen, T, Capelari, M, Nakasone, K, Tewari, JP, Miller, AN, Decock, C, Huhndorf, S, Wach, M, Vishniac, HS, Yohalem, DS, Smith, D, Glenn, AE, Spiering, M, Lindner, DL, Schoch, C, Redhead, SA, Ivors, K, Jeffers, SN, Geml, J, Okafor, F, Spiegel, FW, Dewsbury, D, Carroll, J, Porter, TM, Pashley, C, Carpenter, SE, Abad, G, Voigt, K, Arenz, B, Methven, AS, Schechter, S, Vance, P, Mahoney, D, Kang, S, Rheeder, JP, Mehl, J, Greif, M, Ngala, GN, Ammirati, J, Kawasaki, M, Fang, YG, Matsumoto, T, Smith, D, Koenig, G, Luoma, D, May, T, Leonardi, M, Sigler, D, Taylor, DL, Gibson, C, Sharpton, T, Hawksworth, DL, Dianese, JC, Trudell, SA, Paulus, B, Padamsee, M, Callac, P, Lima, N, White, M, Barreau, C, Juncai, MA, Buyck, B, Rabeler, RK, Liles, MR, Estes, D, Carter, R, Herr Jr., JM, Chandler, G, Kerekes, J, Cruse-Sanders, J, Márquez, RG, Horak, E, Fitzsimons, M, Döring, H, Yao, S, Hynson, N, Ryberg, M, Arnold, AE and Hughes, K. (2008) Preserving Accuracy in GenBank. Science 319 (5870): 1616a.

Cannon, PF, Bridge, PD, and Monte, E. 2000. Linking the past, present, and future of *Colletotrichum* systematics. Pp. 1-20 in: Colletotrichum: Host specificity, pathology, and host-pathogen interaction. D. Prusky, S. Freeman and M. B. Dickman (eds.). APS Press, St. Paul, MN.

Crouch, JA, Clarke, BB, and Hillman, BI. 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. J. 10: 186-195.

Crouch, JA, Clarke, BB and Hillman, BI. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology **96**: 46-60.

Crouch, JA, Tredway, LP, Clarke, BB and Hillman, BI. 2008(a). Ecological specialization drives the evolution of the pathogenic fungus *Colletotrichum cereale* and allied taxa across diverse grass communities. In review with Mol. Ecol.

Dettman JR, Jacobson DJ, Taylor JW (2003) A multilocus genealogical approach to phylogenetic species recognition in the model eukaryote *Neurospora*. *Evolution* **57**, 2703-2720.

Du, M, Schardl, CL, Nuckles, EM, and Vaillancourt, L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Colletotrichum* species complexes. Mycologia 97(3):641-58.

Huelsenbeck, JP, and Ronquist, F. 2001. MrBayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754-755.

Moriwaki, J, Tsukiboshi, T, and Sato, T. 2002. Grouping of *Colletotrichum* species in Japan based on rDNA sequences. J. Gen. Plant Pathol. 68:307-320.

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, and Koljalg, U. 2006 Taxonomic Reliability of DNA Sequences in Public Sequence Databases: A Fungal Perspective. PLoS ONE 1(1): e59. doi:10.1371/journal.pone.0000059.

Posada, D, and Crandall, KA. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-818.

Sherriff, C, Whelan, MJ, Arnold, GM, Lafay, JF, Brygoo, Y and Bailey, JA. 1994. Ribosomal DNA sequence analysis reveals new species grupings in the genus *Colletotrichum*. Exp. Mycol. 18: 121-138.

Sherriff, C, Whelan, MJ, Arnold, GM, and Bailey, JA. 1995. rDNA sequence analysis confirms the distinction between *Colletotrichum graminicola* and *C. sublineolum*. Mycol. Res. 99:475-478.

Sreenivasaprasad, S, Mills, PR, Meehan, BM, and Brown, AE. 1996. Phylogeny and systematics of 18 *Colletotrichum* species based on ribosomal DNA spacer sequences. Genome 39:499-512.

Sutton, BC. 1965. Studies on the taxonomy of *Colletotrichum* Cda with especial reference to *C. graminicola* (Ces.) Wilson, University of London, London.

Sutton, BC. 1966. Development of fruitifications in *Colletotrichum graminicola* (Ces.) Wils. and related species. Can. J. Bot. 44:887-897.

Sutton, BC. 1968. The appressoria of *Colletotrichum graminicola* and *C. falcatum*. Can. J. Bot. 46:873-876.

Sutton, BC. 1980. The coelomycetes: fungi imperfecti with pycnidia, acervuli, and stromata. Kew, U.K.: Commonwealth Mycological Institute.

Sutton, BC. 1992. The genus *Glomerella* and its anamorph *Colletotrichum*. Pp. 1-26 *In Colletotrichum*: Biology, pathology and control. J. A. Bailey and M. J. Jeger (eds.). CAB International, Wallingford, U.K.

Swofford, DL. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Taylor, JW, Jacobson, DJ, Kroken, S, Kasuga, T, Geiser, DM, Hibbett, DS, and Fisher, MC. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31:21-32.

Thompson, JD, Higgins, DG, and Gibson, TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. Nucleic Acids Res. 22:4673-4680.

White, TJ, Bruns, T, Lee, S, and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pp. 315-322 *In* PCR Protocols: A Guide to Methods and Applications. M. A. Innis, D. H. Gelfand, J. J. Sninsky and T. J. White (eds.). Academic Press, San Diego.

Table 7.1

The 55 strains of *Colletotrichum* isolated from grass hosts used in this study

Isolate name	Species	Host plant species	Country of origin	Other available origination data	Year	Source	Pop.
279189-AA	*C. axonopodus*	*Axonoponus affinis*	Australia	Queensland	1983	IMI279189	—
176617-BB	*C. caudatum*	*Bothriochloa bladhii*	Australia	—	1973	IMI 176617	—
176619-IC	*C. caudatum*	*Imperata cylindrica*	Australia	Caboolture	1973	IMI 176619	—
305700-IC	*C. caudatum*	*Imperata cylindrica*	Japan			MAFF 305700	—
238575-ZT	*C. caudatum*	*Zoysia tennufolia*	Japan			MAFF 238575	—
24049-AS	*C. cereale*	*Avena sativa*	Germany	—	1949	CBS 240.49	A1
CA-ANCG17-14	*C. cereale*	*Poa annua*	USA	Pasedena, CA	2004	—	A10
305377AE	*C. cereale*	*Arrhenatherum elatius*	Japan	Chiba Prefecture	1967	MAFF 510634	A2
KS-TA-4-F4	*C. cereale*	*Triticum aestivum*	USA	Marshall County, KS	2005	—	A3
305429-PF	*C. cereale*	*Polypogon fugax*	Japan	Saga Prefecture	1977	MAFF 305429	A4
305076-AS	*C. cereale*	*Avena sativa*	Japan	Saga Prefecture	1966	MAFF 305076	A5
IL-CI-7.3D	*C. cereale*	*Calamagrostis inexpansa*	USA	Markham, IL	2005	—	A6
NJ-8626	*C. cereale*	*Poa annua*	USA	Middletown, NJ	2004	—	A7
68188-lg	*C. cereale*	"lawn grass"	Netherlands	—	1988	CBS 681.88	A8
NJ-HF2B	*C. cereale*	*Poa annua*	USA	New Brunswick, NJ	2003	—	A9
NJ-CA1C1	*C. cereale*	*Calamagrostis acutifolia*	USA	Barrington, NJ	2005	—	B
NJ-DG-2A2-5	*C. cereale*	*Dactylis glomerata*	USA	Sussex County, NJ	2005	—	B
511155-EI	*C. eleusines*	*Eleusine indica*	Japan	Kumamoto Prefecture	1977	MAFF 511155	—
78362-SO	*C. falcatum*	*Saccarum officinarum*	Nigeria		1960	IMI 78362	—
16970-SO	*C. falcatum*	*Saccharum officinarum*	Brazil		1970	CBS 169.70	—
24362-SO	*C. falcatum*	*Saccharum officinarum*	Brazil		1962	CBS 243.62	—
305077-SO	*C. falcatum*	*Saccharum officinarum*	Japan	Chiba Prefecture	1966	MAFF 305077	—
306170-SO	*C. falcatum*	*Saccharum officinarum*	Japan			MAFF 306170	—
306299-SO	*C. falcatum*	*Saccharum officinarum*	Japan			MAFF 306299	—
347765-SO	*C. falcatum*	*Saccharum officinarum*	Nigeria		—		—
M1001	*C. graminicola*	*Zea mays*	USA	Missouri	1978	—	—
IN-900190	*C. graminicola*	*Zea mays*	USA	Indiana	1990	—	—
IN-12475	*C. graminicola*	*Zea mays*	USA	Indiana	1975	—	—
MO-178	*C. graminicola*	*Zea mays*	USA	Missouri	1978	—	—
KY-197	*C. graminicola*	*Zea mays*	USA	McClean County, IN	1997	—	—
NY-15182	*C. graminicola*	*Zea mays*	USA	Tioga Co., NY	1982	—	—
27554-ZM	*C. graminicola*	*Zea mays*	Netherlands	—	1954	CBS 275.54	—
311343-ZM	*C. graminicola*	*Zea mays*	Japan	—	1985	MAFF 311343	—
305404-DC	*C. hanaui*	*Digitaria ciliaris*	Japan	Tochigi Prefecture	1975	MAFF 305404	—
511014-DC	*C. hanaui*	*Digitaria ciliaris*	Japan	Tochigi Prefecture	1975	MAFF 511014	—
1040-DS	*C. hanaui*	*Digitaria* sp.	USA	Monticello, AR	1992	CDG 1040	—
1391-DS	*C. hanaui*	*Digitaria* sp.	USA			CDG 1391	—
305439-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Miyazaki Prefecture	1977	MAFF 305439	—
305460-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Tochigi Prefecture	1980	MAFF 305460	—
511152-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Kochi Prefecture	1977	MAFF 511152	—
511328-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Tochigi Prefecture	1980	MAFF 511328	—
511344-EE	*C. jacksonii*	*Echinochloa esculenta*	Japan	Tochigi Prefecture	1985	MAFF 511344	—

Table 7.1, continued

Isolate name	Species	Host plant species	Country of origin	Other available origination data	Year	Source	Pop.
510857-MS	*C. miscanthi*	*Miscanthus sinensis*	Japan	Tochigi Prefecture	1972	MAFF 510857	—
1047-PD	*C. nicholsonii*	*Paspalum dilatatum*	USA	Baldwin Springs, AR	1985	CDG 1047	—
305391-PD	*C. nicholsonii*	*Paspalum dilatatum*	Japan	Chiba Prefecture	1974	MAFF 305391	—
510916-PD	*C. nicholsonii*	*Paspalum dilatatum*	Japan	Chiba Prefecture	1975	MAFF 510916	—
511115-PD	*C. nicholsonii*	*Paspalum dilatatum*	Japan			MAFF 511115	—
305403-PN	*C. nicholsonii*	*Paspalum notatum*	Japan	Yamaguchi Prefecture	1977	MAFF 305403	—
305428-PD	*C. nicholsonii*	*Paspalum notatum*	Japan	Kumamato Prefecture	1977	MAFF 305428	—
511000-PN	*C. nicholsonii*	*Paspalum notatum*	Japan		1975	MAFF 511000	—
5171-PS	*C. nicholsonii*	*Paspalum* sp.	New Zealand	Auckland	1965	ICMP 5171	—
S3001	*C. sublineolum*	*Sorghum bicolor*	Burkina Fasso		—	—	—
305360-SB	*C. sublineolum*	*Sorghum bicolor*	Japan		1957	MAFF 305360	—
510021-SB	*C. sublineolum*	*Sorghum bicolor*	Japan		1957	MAFF 510021	—
TX-BI2K	*C. sublineolum*	*Sorghum halapense*	USA	Brazoria County, TX	2005	—	—

Table 7.2

Colletotrichum isolated from non-graminicolous hosts used in this study (FN group).

Isolate name	Species	Host plant	Location	Location	Year	Source
23699	*C. dematium*	Unknown	Japan	--	--	MAFF 23699
237705	*C. dematium*	Unknown	Japan	--	--	MAFF 237705
840067	*C. dematium*	Unknown	Japan	--	--	MAFF 840067
840068	*C. dematium*	Unknown	Japan	--	--	MAFF 840868
11-AA	*C. dematium*	*Aeschynomene americana*	USA	Fort Pierce, FL	1982	CDG 11
74-AS	*C. dematium*	*Amaranthus* sp.	USA	Sumter Co., AR	1983	CDG 74
851-CS	*C. dematium*	*Crotalaria spectabilis*	USA	Gainesville, FL	1988	CDG 851
202-PP	*C. dematium*	*Polygonum pennsylvanicum*	USA	Fayetteville, AR	1984	CDG 202
1072-PL	*C. dematium*	*Pueraria lobata*	USA	Lake Weddington, AR	1990	CDG 1072
1075-PL	*C. dematium*	*Pueraria lobata*	USA	Washington Co., AR	1990	CDG 1075
938-PL	*C. dematium*	*Pueraria lobata*	USA	Washington Co., AR	1971	CDG 938
680-TV	*C. dematium*	*Trillium viridens*	USA	--	1986	CDG 680
682-TV	*C. dematium*	*Trillium viridens*	USA	--	1986	CDG 682
305982-PE	*C. dematium*	*Passiflora edulis*	Japan	--	--	MAFF 305982
1349-MS	*C. trifolli*	*Medicago sativa*	New Zealand	--	1992	CDG 1349
1456-GM	*C. truncatum*	*Glycine max*	USA	Fayetteville, AR	1993	CDG 1456

Table 7.3

Species identity of the *Colletotrichum* strains from this as determined through searches of GenBank using ITS sequences.

Isolate name	Species	Species identification using BLAST search with ITS	Best BLAST matches
279189-AA	*C. axononopus*	*C. sublineolum?*	*C. sublineolum, C. dematium, C. truncatum, C. graminicola, C. cereale*
176617-BB	*C. caudatum*	*C. caudatum*	*C. caudatum, C. falcatum, C. cereale, C. brassicae, C. graminicola*
176619-IC	*C. caudatum*	*C. caudatum*	*C. caudatum, C. cereale, C. graminicola*
238575-ZT	*C. caudatum*	*C. caudatum*	*C. caudatum, C. graminicola, C. cereale*
24049-AS	*C. cereale*	*C. graminicola*	*C. graminicola, C. brassicae, C. truncatum, C. dematium, C. cereale, C. coccodes*
CA-ANCG17-14	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. dematium*
305377AE	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. brassicae, C. truncatum, C. dematium*
KS-TA-4-F4	*C. cereale*	*C. dematium or C. truncatum*	*C. dematium, C. truncatum, C. sublineolum, C. cccodes, C. gloeosporioides, C. circnans, C. capsici, C. graminicola, C. cereale*
305429-PF	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. brassicae, C. truncatum, C. capsici, C. dematium*
305076-AS	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. falcatum, C. caudatum, C. brassicae, C. dematium C. capsici*
IL-CI-7.3D	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. brassicae, C. dematium, C. capsici*
NJ-8626	*C. cereale*	*C. cereale or C. graminicola*	*C. graminicola, C. cereale, C. brassicae, C. dematium, C. truncatum*
68188-lg	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. dematium, C. capsici*
NJ-HF2B	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. truncatum, C. dematium, C. capsici, C. brassicae*
NJ-CA1C1	*C. cereale*	*C. graminicola*	*C. graminicola, C. cereale, C. dematium, C. trucatum, C. brassicae*
NJ-DG-2A2-5	*C. cereale*	*C. dematium or C. truncatum*	*C. dematium, C. truncatum, C. capsici, C. circinans, C. graminicola, C. cereale*
511155-EI	*C. eleusines*	*C. graminicola*	*C. graminicola, C. falcatum, C. cereale, C. caudatum*
78362-SO	*C. falcatum*	*C. falcatum*	*C. falcatum, C. graminicola, C. cereale*
16970-SO	*C. falcatum*	*C. graminicola*	*C. graminicola, C. cereae, C. brassicae, C. truncata, C. dematium*
24362-SO	*C. falcatum*	*C. sublineolum*	*C. sublineolum, C. dematium, C. truncatum, C. brassicae, C. capsici, C. coccodes, C. gloeosporioides*
305077-SO	*C. falcatum*	*C. graminicola*	*C. graminicola, C. cereale, C. sublineolum, C. caudatum, C. truncatum, C. falcatum*
306170-SO	*C. falcatum*	*C. falcatum*	*C. falcatum, C. cereale, C. caudatum, C. graminicola*
306299-SO	*C. falcatum*	*C. falcatum*	*C. falcatum, C. cereale, C. caudatum, C. graminicola*
347765-SO	*C. falcatum*	*C. falcatum*	*C. falcatum, C. cereale, C. caudatum, C. graminicola*
M1001	*C. graminicola*	*C. graminicola*	*C. graminicola, C. cereale, C. sublineolum, C. dematium, C. caudatum, C. truncatum*
IN-12475	*C. graminicola*	*C. graminicola*	*C. graminicola, C. dematium, C. cereale, C. capsici*
MO-178	*C. graminicola*	*C. graminicola*	*C. graminicola, C. cereale, C. sublineolum, C. dematium, C. caudatum, C. truncatum*
KY-197	*C. graminicola*	*C. graminicola*	*C. graminicola, C. sublineolum, C. cereale, C. coccodes*
NY-15182	*C. graminicola*	*C. graminicola*	*C. graminicola, C. cereale, C. brassicae, C. dematium, C. capsici*
311343-ZM	*C. graminicola*	*C. graminicola*	*C. graminicola, C. brassicae, C. cereale, C. dematium, C. truncatum*
305404-DC	*C. hanau*	*C. graminicola*	*C. graminicola, C. caudatum, C. brassicae, C. truncatum, C. cereale, C. dematium*
511014-DC	*C. hanau*	*C. graminicola*	*C. graminicola, C. cereale, C. dematium, C. brassicae, C. truncatum*
1040-DS	*C. hanaui*	*C. graminicola* or *C. cereale*	*C. graminicola, C. cereale, C. capsici, C. dematium, C. gloeosporioides,C. coccodes, C. sublineolum.*
1391-DS	*C. hanaui*	*C. fuscum?*	*C. fuscum, C. destructivum, C. coccodes.*
305439-EE	*C. jacksonii*	*C. falcatum*	*C. falcatum, C. graminicola, C. cereale*
305460-EE	*C. jacksonii*	*C. falcatum*	*C. falcatum, C. graminicola, C. cereale*

Table 7.3, continued

Isolate name	Species	Species identification using BLAST search with ITS	Best BLAST matches
511152-EE	*C. jacksonii*	*C. graminicola*	*C. graminicola, C. falcatum, C. cereale, C. caudatum*
511328-EE	*C. jacksonii*	*C. graminicola*	*C. graminicola, C. falcatum, C. cereale, C. caudatum*
511344-EE	*C. jacksonii*	*C. graminicola*	*C. graminicola, C. falcatum, C. cereale, C. caudatum*
510857-MS	*C. miscanthi*	*Unknown*	*C. cereale, C. graminicola, C. dematium, G, falcatum*
305391-PD	*C. nicholsonii*	*C. graminicola*	*C. cereale, C. graminicola, C. caudatum, C. falcatum*
510916-PD	*C. nicholsonii*	*C. graminicola*	*C . graminicola, C. cereale, C. caudatum, C. falcatum*
511115-PD	*C. nicholsonii*	*C. graminicola*	*C . graminicola, C. cereale, C. caudatum, C. falcatum*
305403-PN	*C. nicholsonii*	*C. caudatum*	*C. caudatum, C. cereale, C. falcatum, C. graminicola*
305428-PD	*C. nicholsonii*	*C. graminicola*	*C. graminicola, C. cereale, C. falcatum*
511000-PN	*C. nicholsonii*	*C. graminicola*	*C. graminicola, C. cereale, C. falcatum*
5171-PS	*C. nicholsonii*	*C. graminicola*	*C. graminicola, C. cereale, C. truncatum, C. dematium, C. brassicae*
S3001	*C. sublineolum*	*C. sublineolum*	*C. graminicola, C. cereale, C. sublineolum, C. falcatum, C. caudatum, C. truncatum*
305360-SB	*C. sublineolum*	*C. sublineolum*	*C. graminicola, C. cereale, C. sublineolum, C. truncatum, C. falcatum, C. caudatum*
510021-SB	*C. sublineolum*	*C. sublineolum*	*C. graminicola, C. cereale, C. sublineolum, C. truncatum, C. falcatum, C. caudatum*
TX-BI2K	*C. sublineolum*	*C. sublineolum* or *C. graminicola*	*C. graminicola, C. sublineolum, C. cereale, C. caudatum, C. truncatum*
1072-PL	*C. dematium*	*C. dematium*	*C. dematium, C. capsici, C. graminicola*
1075-PL	*C. dematium*	*C. dematium or C. truncatum*	*C. truncatum, C. dematium, C. capsici, C. brassicae, C.graminicola*
11-AA	*C. dematium*	*C. capsici*	*C. capsici, C. dematium, C. truncatum, C. gloeosporioides, C. boninense*
1349-MS	*C. trifolli*	*C. dematium*	*C. dematium, C. capsici, C. truncatum, C. graminicola*
1456-GM	*C. truncatum*	*C. capsici*	*C. capsici, C. truncatum, C. dematium, C. gloeosporioide*
202-PP	*C. dematium*	*C. truncatum or C. dematium*	*C. truncatum, C. dematium, C. brassicae, C. circinans*
23699	*C. dematium*	*C. coccodes*	*C. coccodes, C. gloeosporioides, C. truncatum, C. dematium, C. sublineolum*
237705	*C. dematium*	*C. dematium or C. truncatum*	*C. truncatum, C. dematium, C. sublineolum, C. brassicae, C. capsici, C. coccodes*
305982-PE	*C. dematium*	*C. truncatum*	*C. truncatum, C. dematium, C. gloeosporioides, C. coccodes*
680-TV	*C. dematium*	*C. dematium or C. truncatum*	*C. sublineolum, C. dematium, C. truncatum, C. capsici, C. graminicola, C,. fuscum*
682-TV	*C. dematium*	*C. dematium or C. truncatum*	*C. sublineolum, C. dematium, C. truncatum, C. capsici, C. graminicola, C,. fuscum*
840068	*C. dematium*	*C. cocoddes*	*C. coccodes, C. gloeosporioides, C. truncatum, C. dematium, C. capsici*
850-CS	*C. dematium*	*C. capsici*	*C. capsici, C, truncatum, C. dematium, C. gloeosporioides*
851-CS	*C. dematium*	*C. capsici*	*C. capsici, C. graminicola, C. dematium, C. truncatum*
938-PL	*C. dematium*	*C. capsici*	*C. capsici, C. gramincola, C. dematium, C. truncatum, C. gloeosporioides*

Table 7.4

Representation and characteristics of *Colletotrichum* and *Glomerella* ITS sequences curated by GenBank. Search of database performed 20-Mar-2008.

	Colletotrichum		***Glomerella***		**Total**	**% of total**
	#	**% of total**	**#**	**% of total**		
Total core nucleotide sequences	1433 [a]		1752 [a]		3185	
ITS sequences	505	35% (of all *Colletotrichum* sequences)	939	54% (of all *Glomerella* sequences)	1444	45%
Unpublished [b]	199	39% (of all *Colletotrichum* ITS sequences)	475	50% (of all *Glomerella* ITS sequences)	674	47%
Published [b]	306	61% (of all *Colletotrichum* ITS sequences)	469	50% (of all *Glomerella* ITS sequences)	775	53%
Most represented species (number of ITS sequences)	*C. cereale* (91) *C. capsici* (52) *C. boninese* (35) *C. dematium* (22) *C. destructivum* (21)		*G. cingulata/C. gloeosporioides* (440) *G. acutata* (350) *G. graminicola/C. graminicola* (63) *G. truncata/C. truncatum* (31)			
Most represented species (number of total sequences)	*G. cingulata/C. gloeosporioides* (1149) *G. acutata/C. acutatum* (596) *C. cereale* (547) *G. grminicola/ C. graminicola* (225)					

[a] 45% of all GenBank sequences are from the anamorphic *Colletotrichum* for which no teleomorph has been described; 55% are from those taxa with a described teleomorphic *Glomerella* state (although the *Colletotrichum* anamorph may have been the only morphological state present)

[b] Publication status refers to whether the ITS sequence curated by GenBank is associated with a peer-reviewed publication and/or a voucher specimen ("published") or is unassociated ("unpublished").

Table 7.5

Reports of new plant disease caused by species of *Colletotrichum* reported in the journal *Plant Disease* 1997-2007, and the methods used for identification.

Host	Species	ITS	Morphology	Citation	Year
Kalmia sp.	*C. acutatum*		X	First Report of *Colletotrichum acutatum* on *Kalmia.* S. R. H. Langrell and S. J. Irvine. 85(4) 447	2001
Alfalfa	*C. truncatum*		X	First Report of *Colletotrichum truncatum* on Alfalfa in Turkey. C. Eken and E. Demirci 84(1): 100.	2000
Strawberry	*C. acutatum*	X	X	First Report of *Colletotrichum acutatum* on Strawberry in Bulgaria. S. G. Bobev, A. Zveibil, and S. Freeman 86(12): 1178	2002
Russian thistle	*C. gloeosporioides*		X	First Report of *Colletotrichum gloeosporioides* on Russian-thistle. I. Schwarczinger, L. Vajna, and W. L. Bruckart 82(12) 1405	1998
Chinese rose	*C. gloeosporioides*		X	First Report of *Colletotrichum gloeosporioides* on Chinese Rose in Argentina. M. C. Rivera, E. R. Wright, and S. Carballo 84(12): 1345	2000
Soybean	*C. coccodes*		X	First Report of *Colletotrichum coccodes* on Soybean in the United States. L. Riccioni, G. Conca, and G. L. Hartman. 82(8): 959	1998
Strawberry	*C. gloeosporioides*		X	First Report of *Colletotrichum gloeosporioides* on Strawberry in Northwestern Argentina M. E. Mónaco, S. M. Salazar, A. Aprea, J. C. Díaz Ricci, J. C. Zembo, and A. Castagnaro 84(5): 595	2000
Draceana sanderiana	*C. draceanaenophilum*			First Report of *Colletotrichum dracaenophilum* on *Dracaena sanderiana* in Bulgaria, S. G. Bobev, L. A. Castlebury, and A. Y. Rossman. 92(1): 173	2008
Blueberry	*C. acutatum*		X	First Report of *Colletotrichum acutatum* in Blueberry Plants in Spain. C. Barrau, B. de los Santos, and F. Romero 85(12): 1285	2001
Tomato	*C. dematium*		X	First Report of *Colletotrichum dematium* on Tomato in Argentina. G. M. Dal Bello. 84(2): 198	2000
Strawberry	*C. acutatum*		X	First Report of *Colletotrichum acutatum* on Strawberry in Northwestern Argentina C. J. Ramallo, L. D. Ploper, M. Ontivero, M. P. Filippone, A. Castagnaro, and J. Díaz Ricci. 84(6): 706	2000
Strawberry	*C. acutatum*		X	First Report of *Colletotrichum acutatum* in Strawberry in Norway. A. Stensvand, G. M. Strømeng, R. Langnes, L. G. Hjeljord, and A. Tronsmo. 85, (5): 558	2001
Cagaita	*C. gloeosporioides*		X	First Report of Anthracnose of Cagaita Caused by *Colletotrichum gloeosporioides* in Brazil. J. R. N. Anjos and M. J. A. Charchar. 85(70): 801	2001
Strawberry	*C. acutatum*		X	First Report of Anthracnose Fruit Rot of Strawberry Caused by *Colletotrichum acutatum* in China.F. M. Dai, X. J. Ren, and J. P. Lu. 90(11): 1460	2006

Table 7.5, continued

Host	Species	ITS	Morphology	Citation	Year
Trichosanthes kirilowii	*C. gloeosporioides*	X	X	First Report of *Colletotrichum gloeosporioides* Causing Anthracnose Fruit Rot of *Trichosanthes kirilowii* in China. H. Y. Li and Z. F. Zhang. 91(5): 636	2007
Tomato	*C. acutatum*	X	X	First Report of Anthracnose Fruit Rot Caused by *Colletotrichum acutatum* on Pepper and Tomato in Bulgaria. Z. J. Jelev, S. G. Bobev, D. Minz, M. Maymon, and S. Freeman. 92(1): 172	2008
Cyclamen persicum	*C. gloeosporioides*		X	First Report of *Colletotrichum gloeosporioides* on *Cyclamen persicum* in Florida D. J. Norman. 81(2): 227	1997
Strawberry	*C. acutatum*		X	First Report of *Colletotrichum acutatum* on Strawberry in Finland. P. Parikka and M. Kokkola. 85(8): 923	2001
Bell pepper	*C. capsici*		X	First Report of *Colletotrichum capsici* on Bell Pepper in Mississippi. K. W. Roy, J. F. Killebrew, and S. Ratnayake. 81(6): 693	1997
Curly dock	*C. destrructivum*		X	First Report of *Colletotrichum destructivum* on Curly Dock. H. B. Lee and C.-J. Kim. 86(11): 1271	2002
Cassava	*C. gloeosporioides*		X	First Report of *Colletotrichum gloeosporioides* f. sp. manihotis, Cause of Cassava Anthracnose Disease, Being Seed-borne and Seed-Transmitted in Cassava. C. N. Fokunang, T. Ikotun, A. G. O. Dixon, and C. N. Akem. 81(6): 695	1997
Gaultheria procumbens	*C. gloeosporioides*	X	X	First Report of Anthracnose of *Gaultheria procumbens* Caused by *Colletotrichum gloeosporioides*. J. F. Elmhirst and E. J. Hudgins. 87(6): 751	2003
Strawberry	*C. acutatum*	X	X	First Report of Anthracnose Caused by *Colletotrichum acutatum* on Strawberry in the Czech Republic. D. Novotn, I. Kíková, J. Krátká, and J. Salava. 91(11): 1516	2007
Taxus mairei	*C. gloeosporioides*	X	X	First Report of Anthracnose Caused by *Colletotrichum gloeosporioides* on *Taxus mairei* in Taiwan. C. H. Fu, W. W. Hsiao, and J. C. Yao. 87(7): 873	2003
Crupina vulgaris	*C. gloeosporioides*	X	X	First Report of Anthracnose of *Crupina vulgaris* Caused by a *Colletotrichum* sp. in Greece. D. K. Berner, F. M. Eskandari, A. Y. Rossman, M. C. Aime, and J. Kashefi. 88(10): 1161	2004
Althaea officinalis	*C. orbiculare*		X	First Report of Anthracnose Caused by *Colletotrichum orbiculare f.* sp. from Marsh Mallow (*Althaea officinalis*) in Switzerland. V. V. Michel. 89(6): 687	2005

Table 7.5, continued

Host	Species	ITS	Morphology	Citation	Year
St. John's Wort	*C. gloeosporioides*		X	First Report of St. John's-Wort Anthracnose Caused by *Colletotrichum gloeosporioides* in Switzerland. N. Debrunner, A.-L. Rauber, A. Schwarz, and V. V. Michel. 84(2): 203	2000
Lygodium microphyllum	*C. gloeosporioides*		X	First Report of the Pathogenicity of *Colletotrichum gloeosporioides* on Invasive Ferns, *Lygodium microphyllum* and *L. japonicum*, in Florida. K. A. Jones, M. B. Rayamajhi, P. D. Pratt, and T. K. Van. 87(1): 101	2003
Althaea officinalis	*C. malvarum*		X	Occurrence of Anthracnose Caused by *Colletotrichum malvarum* on *Althaea officinalis* in Italy. L. Tosi, R. Buonaurio, and C. Cappelli. 88(4): 425	2004
Cowpea	*C. dematium*		X	*Colletotrichum dematium*: Causal Agent of a New Cowpea Stem Disease in South Africa. J. E. Smith and T. A. S. Aveling. 81(7): 832	1997
Pitahaya	*C. gloeosporioides*	X	X	First Occurrence of Anthracnose Caused by *Colletotrichum gloeosporioides* on Pitahaya. A. J. Palmateer, R. C. Ploetz, E. van Santen, and J. C. Correll. 91(5): 631	2007
Avocado	*C. gloeosporioides*	X	X	First Report of the Anamorph of *Glomerella acutata* Causing Anthracnose on Avocado Fruits in Mexico. G. Avila-Quezada, H. V. Silva-Rojas, and D. Teliz-Ortiz. 91(9) 1200	2007
Potato	*C. coccodes*		X	Occurrence of Black Dot of Potato Caused by *Colletotrichum coccodes* in Central Italy. R. Buonaurio, G. Natalini, L. Covarelli, and C. Cappelli. 86(5): 562	2002
Strawberry	*C. acutatum*		X	Occurrence of *Colletotrichum acutatum*, Causal Organism of Strawberry Anthracnose in Southwestern Spain. B. de los Santos Ga de Paredes and F. Romero Muñoz. 83(3): 301	1999
Passion fruit	*C. gloeosporioides*		X	First Report of Anthracnose Caused by *Glomerella cingulata* on Passion Fruit in Argentina. S. Wolcan and S. Larran. 84(6): 706	2000
Common guava	*C. gloeosporioides*		X	First Report of *Glomerella cingulata* on Common Guava in Argentina. M. Carranza, S. Larran, and B. Ronco 86(4): 440	2002
Dogwood	*C. acutatum*	X	X	A New Disease of Flowering Dogwood Caused by *Colletotrichum acutatum*. J. O. Strandberg. 85(2): 229	2001
Azalea japonica	*C. acutatum*		X	Presence of *Colletotrichum acutatum* Causing Leaf Spot on *Azalea japonica* in Italy.A. Garibaldi, D. Bertetti, O. Vinnere, and M. L. Gullino. 88(5): 572	2004
Myrica cerifera	*C. acutatum*	X	X	First Report of *Colletotrichum acutatum* Causing a Leaf Spot on *Myrica cerifera* in Florida. S. J. Mackenzie, L. M. Takahashi, J. C. Mertely, T. E. Seijo, and N. A. Peres. 90(9): 1263	2006

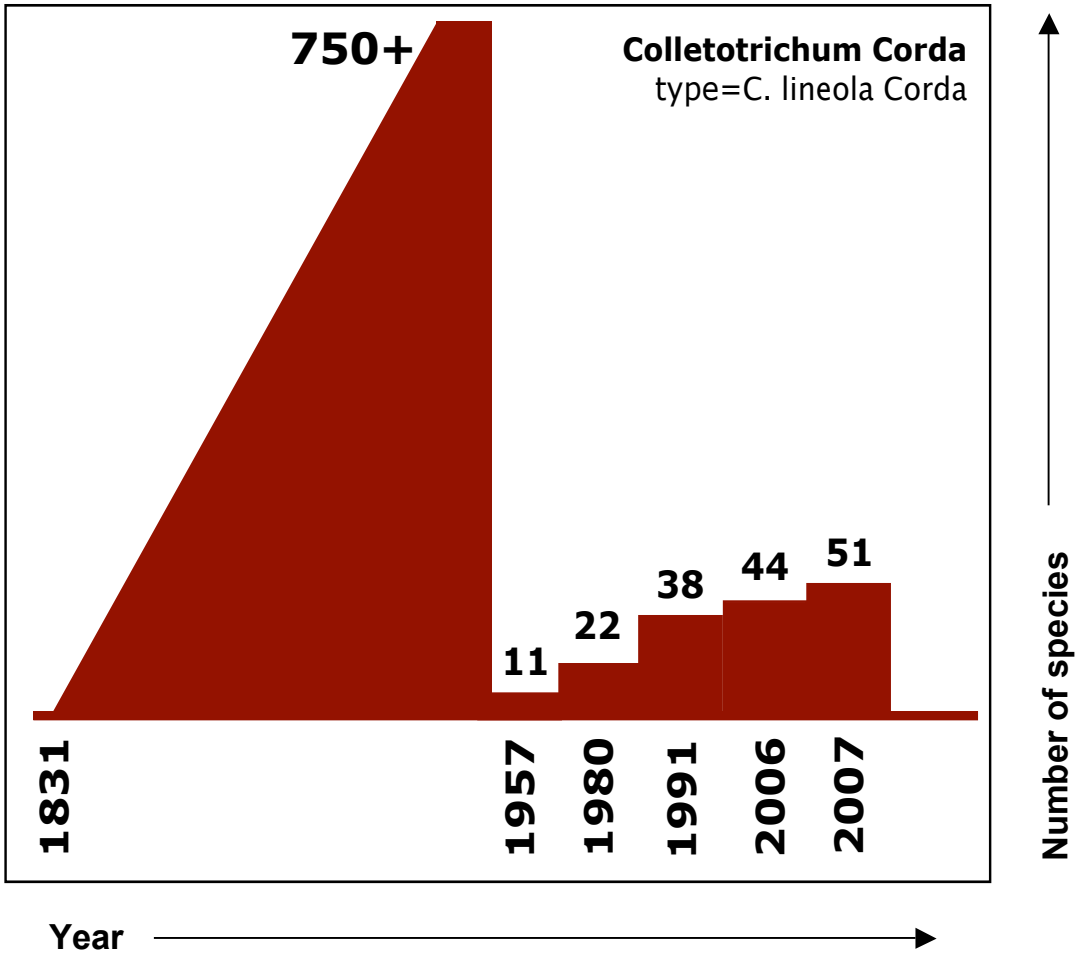


Figure 7.1
The *Colletotrichum* taxonomy timeline (not to scale).
Prior to 1957, new species are described for every *Colletotrichum* identified from a novel plant host. 1957 ushers in the "neo-*Colletotrichum*" era where the treatment of vonArx reduces the number of accpeted species from >750 to 11. Between 1962 to 1991 -- "The Suttonian synthesis" dominated *Colletotrichum* taxonomy, as Sutton resolved some of the problems and generalizations generated by the work of vonArx. 1994 began the "era of the ITS sequence" that continues through this day (2008), where ITS sequence and gene trees dominate the field. But multilocus molecular phylogenetics are beginning to yield real data that is improving our understanding of evolutionary relationships and species boundaries in this important fungal genus.

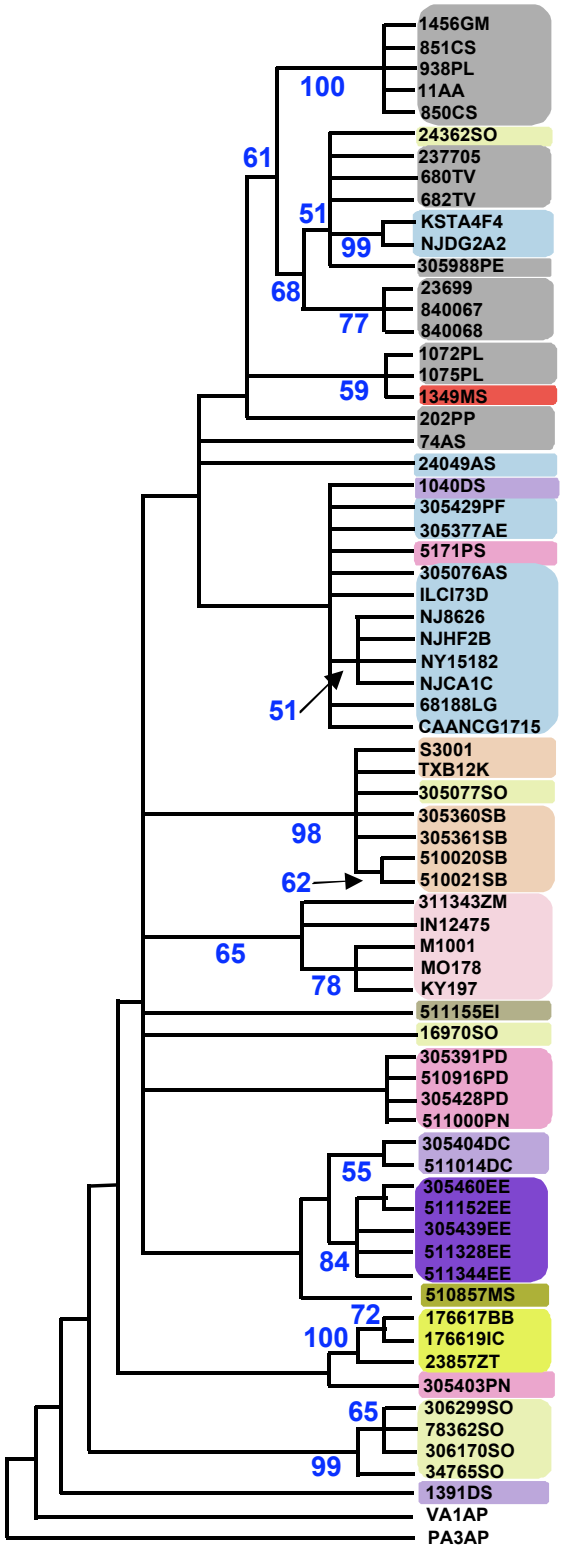


Figure 7.2
Strict consensus tree from the ITS datset analyzed through maximum likelihood. Posterior probabilities >50 supporting the nodes are shown.

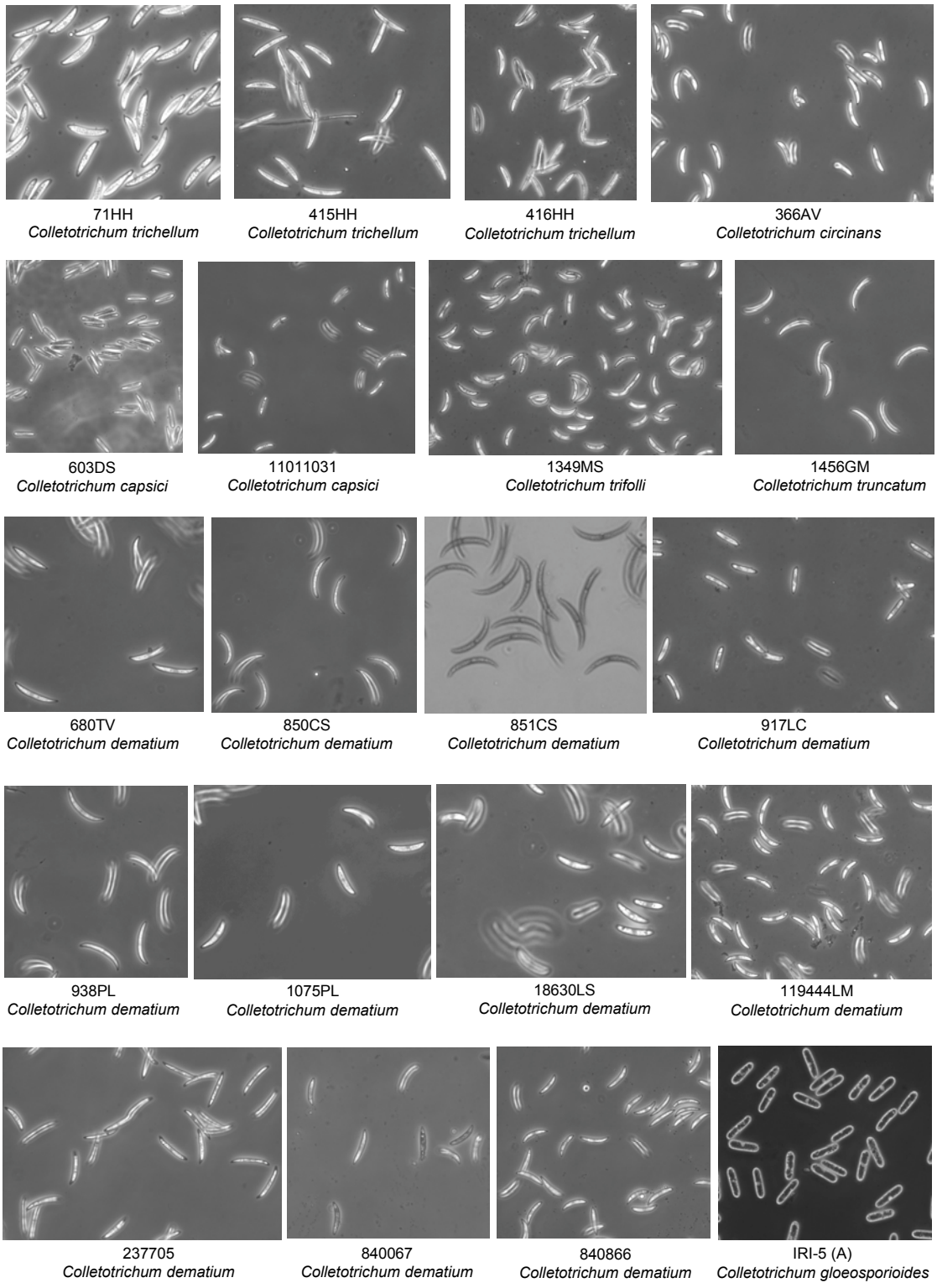


Figure 7.3

Falcate shaped conidia of *Colletotrichum* isolated from non-graminicolous host plants. The photograph in the lower right corner shows an oval-spored species (*C. gloeosporioides*) for comparison. Bar=25 μm.

Appendix 1 GenBank accession numbers of nucleotide sequence data generated by the research described in this dissertation, as of 31-March-2008

Sequences can be accessed through the National Center for Biotechnology Information (NBCI) GenBank database at http://www.ncbi.nlm.nih.gov

Chapter 1 (318 sequences, three genes):
DQ126157–DQ126262, DQ131924 –DQ132027; DQ132028–DQ132051 and DQ133257–DQ133340

Chapter 2 (159 sequences, multiple genes):
DQ663091–DQ663113, DQ663495–DQ663534, DQ666106–DQ666136, DQ666147–DQ666164, DQ667983–DQ668029 and EF067890–EF067895

Chapter 3 (20 sequences, three genes):
DQ663514–DQ663534

Chapter 4 (323 sequences as of 03/19/08, five genes):
EU364976–EU365125, EU554068–EU554157, EU554177–EU554262.

Chapter 5 (31 sequences):
EU554365–EU554396

Appendix 2: Publication version of Chapter 1

Mycology

e-Xtra*

Unraveling Evolutionary Relationships Among the Divergent Lineages of *Colletotrichum* Causing Anthracnose Disease in Turfgrass and Corn

Jo Anne Crouch, Bruce B. Clarke, and Bradley I. Hillman

Department of Plant Biology and Pathology, Rutgers University, 59 Dudley Road, New Brunswick, NJ 08901.
Accepted for publication 30 August 2005.

ABSTRACT

Crouch, J. A., Clarke, B. B., and Hillman, B. I. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology 96: 46-60.

Colletotrichum species cause anthracnose diseases on a number of grass hosts and are common inhabitants of many others. They are divided into four species: *C. sublineolum* is pathogenic to *Sorghum* spp.; *C. caudatum* is found on C4 grasses such as indiangrass and big bluestem; *C. falcatum* causes red rot of sugarcane; and *C. graminicola* sensu lato is a broadly defined species including isolates that attack maize, wheat, oats, and many forage, turf, and amenity grasses of the subfamily Pooideae. In this paper, a combination of hierarchal- and nonhierarchal-based analyses were employed to examine evolutionary relationships among the grass-infecting *Colletotrichum* species, with special emphasis on isolates from turf and other grasses in the subfamily Pooideae. Reconstructions performed with data sets from over 100 *Colletotrichum* isolates at three variable loci using phylogenetic and network-based methodologies unambiguously supported the taxonomic separation of maize-infecting isolates of *C. graminicola* from the pooid-infecting strains of *Colletotrichum*. To reflect the evolutionary relationships that exist between these distinct lineages, we propose the resurrection of the species name *C. cereale* to describe the pooid-infecting isolates. There was also support for further subdivision of *C. cereale*, but the current data are insufficient to confidently subdivide the species, as there was some evidence of recombination between lineages of this species.

When confronted with fungi that rapidly emerge as destructive pathogens in cultivated plant communities, an understanding of how genetic variation is organized often allows us to reconstruct the sequence of events influencing both the onset and trajectory of disease epidemics. By considering fungal phytopathogen life histories and diversity, it has been possible to deduce the events that have shaped many of the most destructive plant disease outbreaks during the past century. For example, the manifestation of the Dutch elm disease pandemics caused by *Ophiostoma* spp. is known to be characterized by a series of rapid intercontinental migratory events, introgressive gene transfers resulting in novel hybrid genotypes, and, significantly, by the replacement of one pathogen (*O. ulmi*) by a new, more destructive species (*O. novo-ulmi*) (11 for review). Similarly, by reconstructing phylogenetic relationships among host-specific populations of the rice blast pathogen (*Magnaporthe oryzae*), a recent study suggests that host shifts of the pathogen from non-rice plants to rice crops occurred during the era of rice domestication (17). Pursuit of this knowledge is not merely academic: with enough information, plant pathologists have a greater probability of implementing successful disease control measures.

In cereals and grasses, anthracnose disease is caused by four distinct *Colletotrichum* species, each delimited partly by host specificity, but primarily differentiated by morphological features. *C. sublineolum* Henn. is pathogenic to *Sorghum* spp.; *C. caudatum* Peck is found on a number of C4 grasses such as indiangrass and big bluestem (*Sorghastrum nutans* (L.) Nash and *Andropogon gerardii* Vitman, respectively); and *C. falcatum* Went causes red rot of sugarcane (*Saccharum officinarum* L.). *C. graminicola* (Ces.) G.W. Wilson is generally considered to inhabit a wide range of hosts including corn (*Zea mays* L.), wheat (*Triticum aestivum* L.), oats (*Avena sativa* L.), and many forage, turf, and amenity grasses of the Poaceae subfamily Pooideae (71). Sutton formally proposed in 1980, however, that the species circumscription is properly applied only to *Colletotrichum* on corn; therefore, valid publication of *C. graminicola* sensu stricto Sutton rendered illegitimate the application of the name *C. graminicola* to any *Colletotrichum* from hosts other than corn (61). As no diagnoses of a new species taxon for the fungi distributed on pooid grasses has been established, *C. graminicola* sensu lato G.W. Wilson is still routinely, but inappropriately, employed to describe these strains (10).

Historically, *C. graminicola* sensu lato G.W. Wilson was known as a plant pathogen of minor importance (18,20), and in almost all grasses this is still true. But toward the end of the 20th century, two devastating anthracnose disease epidemics caused by *C. graminicola* occurred quite suddenly in North America: first, in corn crops and later in turfgrasses of the subfamily Pooideae cultivated as golf course greens (especially annual bluegrass [*Poa annua* L.] and creeping bentgrass [*Agrostis stolonifera* L.]). The recent disease upsurge caused by *C. graminicola* has devastated turfgrasses maintained as golf course greens since the early 1990s and shows no signs of abating. In turf, changing cultivation practices are likely a contributing factor to disease development, as stands of grass are exposed to increasingly harsh management regimes designed to enhance playability for golfers. Factors such as minimal nitrogen fertilization, decreased cutting heights, variability in fungicide efficacy, and the increased usage of plant growth regulating chemicals have all been implicated in the enhancement of

Corresponding author: B. I. Hillman; E-mail address: hillman@aesop.rutgers.edu

All sequences have been deposited in GenBank (accession nos. DQ126157 to DQ126262, DQ131924 to DQ132027, DQ132028 to DQ132051, and DQ133257 to DQ133340).

*The *e*-Xtra logo stands for "electronic extra" and indicates that the online version contains supplemental material not included in the print edition. The online version contains maximum parsimony trees of ITS, HMG box, and SOD2 regions, and a representative sample of the sequence alignments that were the basis for those trees.

DOI: 10.1094/PHYTO-96-0046
© 2006 The American Phytopathological Society

Appendix 2: Publication version of Chapter 1

anthracnose disease levels (22; J. Inguagiato and B. B. Clarke, *unpublished data*).

As anthracnose disease in turfgrasses attributed to *C. graminicola* became increasingly problematic throughout the 1990s, a number of research groups conducted analyses of molecular genetic diversity or inferred gene genealogies intended to characterize variability in the *C. graminicola* species group. Several conflicting hypotheses were proposed concerning the evolution and diversity of *C. graminicola*, but no theory has generally been accepted as an accurate representation of the species history. Molecular data were used to suggest a close connection between *C. sublineolum* and isolates of *Colletotrichum* causing anthracnose in annual bluegrass turf (6,30), whereas another study supported an association between isolates from creeping bentgrass turf and those from corn (6). Results from other research groups supported a separate lineage of turfgrass pathogens, diverged from corn and sorghum isolates, which share a single ancestor (12,19,25,31).

In this study, we jointly consider three complementary data sets to investigate the patterns of evolutionary relationships among *Colletotrichum* isolates responsible for anthracnose disease in turfgrasses, corn, and other grass hosts. We tested the following expectations: (i) *C. graminicola* sensu lato G.W. Wilson is actually an assemblage of divergent lineages, rather than a single cohesive species; and (ii) although this fungus appears to reproduce clonally through asexual conidia (16), with the sexual state (*Glomerella*) not yet observed in a natural setting, lineages of these fungi may still be structured to some degree by reticulate evolution. To consider these concepts, data from over 100 *Colletotrichum* isolates from grasses and cereals were collected from more than 88 populations throughout the United States and Canada. A combination of hierarchal- and nonhierarchal-based approaches that were used to test the limits of species boundaries rigorously provides compelling evidence of sympatric, species-level phylogenetic divergence between *C. graminicola* pathogenic to maize and *Colletotrichum* populations causing disease in grasses of the subfamily Pooideae, including cultivated C3 turfgrass species. In particular, these analyses demonstrated that *Colletotrichum* spp. isolated from maize and pooid grasses are each more closely related to *C. sublineolum* from sorghum than these two lineages are to one another. To reflect the evolutionary relationships that exist between these distinct organisms, we formally propose to resurrect and revise the species *C. cereale* Manns (54) to describe the unique lineages isolated from pooid grasses, while *C. graminicola* sensu stricto Sutton uniquely describes *Colletotrichum* from corn hosts. To aid communication throughout this narrative and to emphasize the fundamental distinction to be made between these highly divergent lineages, we will henceforth refer to any *Colletotrichum* specimen isolated from pooid grasses as a member of the *C. cereale* species group, and *C. graminicola* as the species of *Colletotrichum* pathogenic to corn.

MATERIALS AND METHODS

Taxon sampling for molecular analysis. A total of 107 specimens of *C. graminicola* isolated from North American grasses (Table 1) were assessed for this study, with the majority of individuals (76%) isolated from diseased stands of turfgrass maintained as golf course greens. This sample collection includes specimens from 88 localities, and closely reflects the geographic regions where anthracnose disease on golf course greens has been most prevalent in North America; not all known localities were sampled.

Collected fungi were established in pure culture on potato dextrose agar (PDA; Fisher Scientific, Hampton, NH) and then single-spore purified to ensure isogenic lines. Cultures were preserved as dehydrated mycelia on Whatman glass fiber at –20°C. Identification of fungal colonies as *Colletotrichum* was performed using spore and setae morphological characters and later confirmed by performing a BLAST search (3) of the partial ribosomal DNA (rDNA) sequences against the National Center for Biotechnology Information (NCBI) database.

Outgroup taxa. Phylogenetic relationships within the genus *Colletotrichum* are currently ill-defined, making the choice of an appropriate outgroup to root our phylogenetic tree uncertain based upon previously published hypotheses. Traditional *Colletotrichum* systematics relies heavily upon host plant association (13); this criterion has been repeatedly shown to be unreliable for the genus; in some instances several species are capable of infecting a single host plant (37,42) or a single species may be associated with an extremely broad host range (28 for review). Morphological characters in the genus may also be systematically uninformative, and are often inadequate to resolve even interspecific relationships with any level of confidence (62). Likewise, molecular investigations of the group conducted to date fail to infer evolutionary relationships between the *Colletotrichum* taxa robustly. Instead, unresolved, minimally supported topologies from which no significant conclusions can be drawn have been found (45). Because of the ambiguous state of *Colletotrichum* taxonomy, we constructed a phylogenetic tree using the nucleotide sequence from the rDNA internal transcribed spacer (ITS) region to select an appropriate species empirically as outgroup taxa for this study. BLAST searches of the NCBI database GenBank using the ITS sequence from strains MO-1001178, PA-50005, and NJ-6340 identified a number of *Colletotrichum* species likely to be closely related to *C. graminicola*. Twenty-four sequences were analyzed (AY536046, AF272786, AF489567, AJ301951, AJ301922, AF272783, AF272782, AB042305, AB042304, AY376530, AB196301, AJ301954, AF411770, AJ536231, AF534469, AJ301975, AJ301968, and AJ311391) along with a representative selection of *C. graminicola* ingroup taxa. Based upon the resultant 75% consensus tree topology (data not shown), four isolates of *C. sublineolum* from sorghum were included in this study, since this species appeared to be a closely related sister taxon to *C. graminicola*. In addition, two isolates of *C. acutatum* J.H. Simmonds were included as outgroup taxa because they are clearly distinct from *C. graminicola* both morphologically and on the molecular level, yet still show a close enough relationship to allow for unambiguous nucleotide sequence alignments.

DNA isolation, amplification and sequencing. Total genomic DNA was extracted from fungal cultures using a standard phenol/chloroform protocol as described (19). Polymerase chain reaction (PCR) products for the ITS region of the *ITS1*, *5.8S*, and *ITS2* rDNA and the conserved HMG-box of the *MAT1-2* mating-type locus (HMG) were generated using published primer pairs and amplification conditions: ITS4 and ITS5 (70) for the ITS region, NcHMG 1 and 2 (4) or CgHMG 1 and 2 (66) for the HMG sequence. New primer pairs were designed to amplify a 625- or 505-bp portion of a single copy manganese-type superoxide dismutase gene (*SOD2*) (26): SOD625F/R (5′GCCCACAGTACATATTGCCTAAGC3′ and 5′TCATCCCGGGAGCCAGAAAACCT3′) or SOD507F/R (5′ATGGCAGCCTTTCCGTTGAGATAC3′ and 5′AGTTGACATGAAGCCACCTACAGC3′). PCR primers were synthesized by Sigma Genosys (The Woodlands, TX), and all amplifications were performed in a Biometra UNO thermoblock (Whatman Biometra, Goettingen, Germany) in 25-µl reactions using 37.5 ng of genomic DNA, 1 unit of *Taq* DNA polymerase (Promega, Madison, WI) in 10× PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, and 12.5 ng of each primer. *SOD2* products were amplified from genomic DNA using an initial denaturation at 95°C for 5 min followed by 40 cycles of denaturation at 95°C for 1 min, annealing at 53°C for 1 min, and extension at 72°C for 1 min; with a final extension at 72°C for 10 min. Negative controls were included in all amplifications to check for possible contamination. Amplified fragments were visualized on 0.8% agarose gels, and then excised and purified using the Gene Clean III Kit (Qbiogene, Irvine, CA). Nucleotide sequences were

Appendix 2: Publication version of Chapter 1

generated from the forward and reverse strands directly from the amplicon using the corresponding PCR primers and BigDye Terminator cycle-sequencing chemistry (Applied Biosystems, Foster City, CA) on an ABI 3100 Capillary Sequencer following the manufacturer's protocol but using a quarter of the suggested reaction volume.

Nucleotide sequences were assembled and edited using the Lasergene Sequence Analysis Software package (DNASTAR, Inc., Madison, WI). Three hundred eighteen DNA sequences were deposited in GenBank under accession nos. DQ126157 to DQ126262 (ITS), DQ131924 to DQ132027 (HMG), and DQ132028 to DQ132051 and DQ133257 to DQ133340 (*SOD2*). The remaining 15 sequences were not reported because they overlapped data already placed by Du et al. (25) (DQ003109-12, DQ003114, DQ003116-17, DQ002855-59, DQ002826, DQ002862, DQ002865, and DQ002868). Sequences were aligned using the Clustal W algorithm (65) and then manually adjusted in Microsoft Word (Microsoft Corp., Redmond, WA). The alignment of protein coding regions was refined according to amino acid sequences and intron positions. Gaps were removed from the data set, coded as single multi-state characters, and then reintroduced as coded data for the phylogenetic analyses.

TABLE 1. Sources of *Colletotrichum* isolates used in this study

Clade[a]	Haplotype (frequency)	Isolate name	Host species	Origin	Source	Original reference
…	1 (1)	NJ-CACA (outgroup)	*Calamagrostis acutifolia* (*C. acutatum*)	New Jersey	…	This study
B	2 (1)	ALB-99325	*Poa pratensis*	Alberta, Canada	T. Hsiang	This study
A	3 (1)	MA-6722	*Poa annua*	Massachusetts	…	This study
A	4 (1)	CT-6956	*Poa annua*	Connecticut	…	This study
A	5 (22)	CA-62	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-1049	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-1143	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-CL9	*Poa annua*	California	F. Wong	This study
A	5 (22)	CA-SH29	*Poa annua*	California	F. Wong	This study
A	5 (22)	CT-2	*Poa annua*	Connecticut	N. Jackson	(12)
A	5 (22)	IL-P6G	*Poa annua*	Illinois	N. Jackson	(12)
A	5 (22)	IL-PT	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	IL-PV1	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	IL-PV2	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	IL-RCC	*Poa annua*	Illinois	R. Kane	This study
A	5 (22)	NJ-6795	*Poa annua*	New Jersey	…	This study
A	5 (22)	NJ-7284	*Poa annua*	New Jersey	…	This study
A	5 (22)	NY-8422	*Poa annua*	New York	…	This study
A	5 (22)	NY-USGA	*Poa annua*	New York	…	This study
A	5 (22)	ONT-00176	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
A	5 (22)	PA-50014	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-50111	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-50231	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-V1	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	PA-WH3	*Poa annua*	Pennsylvania	W. Uddin	This study
A	5 (22)	RI-8	*Poa annua*	Rhode Island	N. Jackson	(12)
A	6 (1)	NBR-13	*Poa annua*	New Brunswick, Canada	N. Jackson	(12)
A	7 (1)	CA-540	*Poa annua*	California	F. Wong	This study
A	8 (1)	CO-8910	*Poa annua*	Colorado	…	This study
A	9 (4)	KS-20DGU	*Dactylis glomerata*	Kansas	…	This study
A	9 (4)	KS-20DGY	*Dactylis glomerata*	Kansas	…	This study
A	9 (4)	KS-20EVD	*Elymus virginicanus*	Kansas	…	This study
A	9 (4)	KS-20EVM	*Elymus virginicanus*	Kansas	…	This study
A	10 (1)	NJ-CA1	*Calamagrostis acutifolia*	New Jersey	…	This study
A	11 (1)	NJ-DG1	*Dactylis glomerata*	New Jersey	…	This study
A	12 (1)	NJ-8627	*Poa annua*	New Jersey	…	This study
A	13 (1)	KS-10EC1A	*Elymus canadensis*	Kansas	…	This study
A	14(1)	NY-8900	*Poa annua*	New York	…	This study
A	15 (1)	NJ-8626	*Poa annua*	New Jersey	…	This study
A	16 (1)	NJ-6340	*Poa annua*	New Jersey	…	This study
A	17 (2)	CA-EG15	*Poa annua*	California	F. Wong	This study
A	17 (2)	CA-SC44	*Poa annua*	California	F. Wong	This study
A	18 (1)	MA-11	*Poa annua*	Massachusetts	N. Jackson	(12)
A	19 (1)	CT-14	*Poa annua*	Connecticut	N. Jackson	(12)
A	20 (1)	KS-20BIG	*Bromus inermis*	Kansas	…	This study
A	17 (2)	NJ-9582	*Poa annua*	New Jersey	…	This study
A	17 (2)	PA-50183	*Poa annua*	Pennsylvania	W. Uddin	This study
A	22 (1)	CA-SC32	*Poa annua*	California	F. Wong	This study
A	23 (1)	CA-1715	*Poa annua*	California	F. Wong	This study
A	24 (1)	RI-22	*Agrostis stolonifera*	Rhode Island	N. Jackson	(12)
B	25 (1)	PA-50002	*Poa annua*	Pennsylvania	W. Uddin	This study
B	26 (1)	CT-28	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
B	27 (1)	NJ-6607	*Poa annua*	New Jersey	…	This study
B	28 (1)	MA-17	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	29 (1)	ONT-00133	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
B	30 (1)	CT-25	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
B	31 (3)	MA-20	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)

(*Continued on next page*)

[a] Clades A to D are inferred from multilocus sequence analyses, as shown in Figure 1.

Appendix 2: Publication version of Chapter 1

Phylogenetic analyses. We used both Bayesian maximum likelihood (ML) and maximum parsimony (MP) methods to infer phylogenetic relationships from each gene region and for the combined data set. Models of evolution and ML parameters were estimated individually for each data set using the Akaike information criterion (AIC) (1) as implemented in Modeltest version 3.06 (47). Bayesian analyses were performed using MrBayes version 3.0b4 (34), which allowed the different data partitions to be modeled separately using the evolutionary models and informative priors that best described the data. The optimal model selected under the AIC implemented in Modeltest was specified as prior for each gene. One cold and three incrementally heated Metropolis-coupled Monte Carlo Markov chains were started from random trees and run simultaneously for 14,000,000 generations each in order to sample the phylogenies according to their posterior probabilities. Three replicate runs were performed, and tree topologies and stationarity levels were compared for convergence (34). To determine the point at which stationarity was achieved in each run, log likelihood scores were plotted against generation time; the initial generations sampled before convergence were discarded as burn-in. We sampled trees from every 500 generations across all four independent analyses to calculate posterior probabilities for each branch in the ML tree. Trees sampled from the posterior distribution were imported into

TABLE 1. (*Continued from preceding page*)

Clade[a]	Haplotype (frequency)	Isolate name	Host species	Origin	Source	Original reference
B	31 (3)	MA-21	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	31 (3)	ONT-00128	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
B	32 (2)	ONT-00124	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	(16)
B	32 (2)	ONT-00126	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
B	33 (1)	PA-50621	*Poa annua*	Pennsylvania	W. Uddin	This study
B	34 (1)	NY-19	*Agrostis stolonifera*	New York	N. Jackson	(12)
B	35 (1)	PA-4410	*Poa annua*	Pennsylvania	W. Uddin	This study
B	36 (1)	NJ-6491	*Poa annua*	New Jersey	…	This study
B	37 (2)	NJ-4990	*Poa annua*	New Jersey	…	This study
B	37 (2)	PA-1112	*Poa annua*	Pennsylvania	W. Uddin	This study
B	38 (2)	PA-50005	*Poa annua*	Pennsylvania	W. Uddin	This study
B	38 (2)	PA-50181	*Poa annua*	Pennsylvania	W. Uddin	This study
B	39 (1)	MA-24	*Agrostis stolonifera*	Massachusetts	N. Jackson	(12)
B	40 (1)	PA-50623	*Poa annua*	Pennsylvania	W. Uddin	This study
B	41 (1)	CT-18	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
A	42 (1)	TX-26	*Agrostis stolonifera*	Texas	N. Jackson	(12)
A	43 (1)	VA-8977	*Poa annua*	Virginia	…	This study
A	44 (1)	ONT-99359	*Agrostis stolonifera*	Ontario, Canada	T. Hsiang	This study
A	45 (1)	NH-23	*Agrostis stolonifera*	New Hampshire	N. Jackson	(12)
A	46 (1)	NJ-8467	*Poa annua*	New Jersey	…	This study
A	47 (5)	NJ-7130	*Agrostis stolonifera*	New Jersey	…	This study
A	47 (5)	NJ-8400	*Lolium perenne*	New Jersey	…	This study
A	47 (5)	NJ-RWCC	*Poa annua*	New Jersey	…	This study
A	47 (5)	NJ-HF2B	*Agrostis stolonifera*	New Jersey	…	This study
A	47 (5)	PA-211	*Poa annua*	Pennsylvania	W. Uddin	This study
B	48 (1)	NJ-6553	*Poa annua*	New Jersey	…	This study
B	49 (1)	CT-27	*Agrostis stolonifera*	Connecticut	N. Jackson	(12)
B	50 (1)	NY-16	*Agrostis stolonifera*	New York	N. Jackson	This study
…	51 (1)	NJ-10BB (outgroup)	*Vaccinium corymbosum* (*C. acutatum*)	New Jersey	P. Oudemans	This study
C	52 (1)	S3001	*Sorghum bicolor* (*C. sublineolum*)	Burkina Fasso	L. Vaillancourt	(67)
D	53 (1)	NY-15182	*Zea mays*	New York	G. Bergstrom	This study
C	54 (2)	S12001	*Sorghum bicolor*	Brazil	L. Vaillancourt	(67)
C	54 (2)	S19001	*Sorghum bicolor*	South Africa	L. Vaillancourt	(25)
C	55 (1)	S17001	*Sorghum bicolor*	Texas	L. Vaillancourt	(25)
D	56 (2)	MO-478	*Zea mays*	Missouri	L. Vaillancourt	This study
D	56 (2)	MO-978	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	BZ-500190	*Zea mays*	Brazil	L. Vaillancourt	(67)
D	57 (18)	IN-10472	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-10570	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-10670	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-10970	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-12270	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-12475	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-300170	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	IN-DUB90	*Zea mays*	Indiana	L. Vaillancourt	This study
D	57 (18)	KY-197	*Zea mays*	Kentucky	L. Vaillancourt	This study
D	57 (18)	KY-297	*Zea mays*	Kentucky	L. Vaillancourt	This study
D	57 (18)	KY-397	*Zea mays*	Kentucky	L. Vaillancourt	This study
D	57 (18)	KY-398	Zea mays	Kentucky	L. Vaillancourt	(25)
D	57 (18)	MO-178	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	MO-878	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	MO-1001178	*Zea mays*	Missouri	L. Vaillancourt	This study
D	57 (18)	NC-200170	*Zea mays*	North Carolina	L. Vaillancourt	This study
D	57 (18)	NY-AU80	*Zea mays*	New York	L. Vaillancourt	(25)
D	57 (18)	NY-400180	*Zea mays*	New York	L. Vaillancourt	This study
D	58 (1)	IN-D77	*Zea mays*	Indiana	L. Vaillancourt	This study
D	59 (1)	IN-900190	*Zea mays*	Indiana	L. Vaillancourt	This study

Appendix 2: Publication version of Chapter 1

PAUP* version 4.0b10 (63) and, after excluding the burn-in, a 75% majority-rule consensus tree was constructed, with the percentage of samples recovering a particular clade representing that clade's posterior probability (33). Bayesian posterior probabilities correspond to the statistical probability that a clade is present in the true tree, given the specified priors, the likelihood model, and the data examined (33,39); a final 75% consensus tree was then generated from all trees in the posterior distribution sampled from the three runs (83,539 trees).

MP analyses were performed using PAUP* under the heuristic search option, with starting trees obtained from 100 random-addition replicates, and using a tree-bisection reconnection (TBR) branch swapping strategy. Multiple MP trees for any data set were combined into single strict consensus trees. Nonparametric bootstrap resampling was conducted to evaluate relative levels of support for individual nodes (27) using 1,000 bootstrap pseudoreplicates with 100 random additions and TBR branch swapping. MP analyses were conducted for both the individual data sets and the total combined data set.

Estimating recombination and its impact on phylogenetic inference. Although *C. graminicola* is thought to reproduce in nature almost exclusively by clonally generated conidia, populations of this fungus still may be structured to some extent by recombination, either through historical events or through the exchange of genetic material by means of vegetative anastomosis between different individuals. Since the presence of recombination has the potential to produce conflicting phylogenetic signal, and may lead to incorrect evolutionary inferences, we investigated whether relationships between the *C. graminicola* taxa are more accurately represented by a reticulating network rather than a bifurcating phylogenetic tree topology. First, topologies of individual gene genealogies and the multilocus gene tree were visually inspected for the presence of incongruence suggestive of recombination. To detect recombination events not reflected in the tree topology, a combination of methods that use substitution patterns and site incompatibility to infer the presence of conflicting signal were employed. After removing indels from the aligned multilocus data set, sequences were collapsed into unique haplotypes using SNAP Map and SITES version 1.1 (29) launched through SNAP Workbench (51). Since the majority of the ingroup taxa in our analyses are likely to be recently diverged, all sites violating the infinite sites model, which assumes that the possibility of multiple mutations at a single nucleotide site can be ignored due to extremely low mutation rates, were eliminated. The resultant haplotype data set was used to identify the presence of reticulating relationships inappropriately represented in an acyclic tree topology; such relationships would be generated by evolutionary processes acting at the population level such as hybridization between lineages or recombination between genes, or homoplasic events such as convergent, parallel, or reverse mutations. We used the split decomposition network method, a transformation-based approach that uses distance parsimony to partition data sets into "splits" of sequences (8). The splits are combined successively, with any incompatible, contradictory groupings introducing a loop into the network to indicate the conflict. We employed the computer program SplitsTree version 4.0 b14 (35) to visualize reticulation networks in the data set, with branch support estimated by performing 10,000 bootstrap pseudoreplicates. In the split-graph network, reticulating taxa will be connected by more than one branch, with each connection representing alternative solutions; if the data are perfectly phylogenetic, a bifurcating tree topology will result.

To explore further the boundary between hierarchal and nonhierarchal, tokogenetic relationships in our sample, a site compatibility matrix was generated from the haplotype data set using the SNAP Clade and SNAP Matrix functions of SNAP Workbench. The compatibility matrix was used to identify the presence of compatibility/incompatibility among the haplotypes, with any resultant incompatible sites removed from the data set. The data set was also evaluated for the signature of recombination events by using Hudson's four gamete test (32) executed in DnaSP (53); any sites identified as potentially recombinant were also excluded from the matrix. The resultant haplotype data set, with any potentially confounding recombinant or homoplasic sites removed, was used to perform MP analysis in PAUP* to visualize those portions of the tree topology uniquely characterized by hierarchal, species-level relationships.

Morphological analysis. To determine if the fungi described in the present study as *C. cereale* conform to the morphological description established by Selby and Manns (54), we inspected syntype specimens from the Kansas State Herbarium, presently curated by the New York Botanical Garden Steere Herbarium (NYBG) (NYBG specimen numbers 305598 [type] and 276683 on *Bromus secalinus*; 305599 on *Avena sativa* [type]; 305595 on *Arrhenatherum elatius* [type]; 305596 [type] and 276680 on *Secale cereale*; 305597 [type], 276684 and 276681 on *Triticum vulgare*; 276686 on *Dactylis glomerata*; 276687 on *Phleum pratense*; 276685 on *Agrostis alba*; and unnumbered samples from *Agrostis alba*, *Arrhenatherum elatius*, and *Poa pratensis*). Morphological examination was also performed on *Colletotrichum* specimens isolated from pooid grasses in our collection (KS-20BIG from *Bromus inermis*; ONT-00128, NY-16, CT-18, MA-20, MA-21, MA-24, CT-25, and CT-27 from *Agrostis stolonifera*; NJ-6722, NJ-6795, CA-1715, NJ-4990, NH-23, NBR-13, RI-9, CT-8, MA-11, PA-50114, PA-WH3, PA-50623, PA-50231, PA-5005, PA-50111, PA-50183, PA-4410, PA-V1, and PA-50002 from *Poa annua*; KS-20DGU and KS-20DGY from *D. glomerata*; NJ-CA1 from *Calamagrostis acutifolia*; and KS-20EVM from *Elymus virginicanus*), with cultures grown on PDA under constant light at 26°C for 7 to 10 days. Microscopic observations and measurements were made with an Olympus CX40 microscope with bright field or phase contrast illumination. At least 50 conidia from each isolate were measured; dimensions of other fungal structures are given as the range of at least 20 measurements where possible. Photomicrographs were obtained using an Olympus BX41 microscope equipped with a MagnaFire Digital Firewire camera and software (Optronics, Goleta, CA).

RESULTS

Data set characteristics and analysis. Because ribosomal DNA genes are maintained at high copy numbers in eukaryotic genomes, nonorthologous copies may potentially be present (46), violating the expectation of homology required to reconstruct phylogenetic relationships accurately. In the present study, two lines of evidence support the presence of only orthologous rDNA sequences: nucleotide base calling from the ITS rDNA chromatographs generated by sequencing directly from the PCR products was clean and unambiguous, and all three gene trees were concordant at main branch points and tip clades where phylogenetic signal is adequate to resolve such relationships.

Among members of the *C. graminicola*/*C. cereale*/*C. sublineolum* ingroup, few insertion-deletion events were observed in the nucleotide sequence alignments, and none of the gaps introduced rendered the ingroup problematic; ambiguous alignment was restricted to outgroup comparisons. The combined molecular data set comprised a total of 1,229 nucleotides, with 130 indels coded as standard characters (Table 2). Much of the coded indel data were autapomorphic only with respect to the outgroup taxa and were otherwise parsimony-informative between ingroup taxa; 8 indels of 2 to 4 bp in length within introns were characteristic of particular monophyletic groupings recovered by the ingroup phylogeny.

Given the number of variable characters in the data sets, all three genes appeared to be potentially informative. A preliminary analysis of a 650-bp region of the *TUB-2* gene from the taxa NJ-

Appendix 2: Publication version of Chapter 1

6340, PA-50005, NJ-6491, MO-1001178, and NY-15182 showed only one variable character; therefore, it was not included in the phylogenetic reconstruction (data not shown). In particular, the 536-bp *SOD2* sequence consists of 41% parsimony-informative characters within the ingroup taxa (ITS = 12%; HMG = 12%; combined = 25%). Individual gene genealogies constructed using strict consensus trees under parsimony showed visually concordant topologies, although the HMG and ITS sequences were unable to recover some of the groups reconstructed by the more informative *SOD2* and the combined data set due to their lower levels of phylogenetic signal. This relative lack of phylogenetic resolution from the *C. graminicola* ITS sequence is consistent with our previously published results (19) and a study conducted by Hsiang and Goodwin (31); similar reports have been made in other *Colletotrichum* species (7). The lower resolution exhibited by the ITS gene tree may also result to some degree from the extremely large number of equally parsimonious tree topologies recovered during the heuristic searches: 44,034 MP trees were inferred from the ITS sequence data (HMG = 4; *SOD2* = 72).

To assess whether the ITS, HMG, and *SOD2* sequences generate tree topologies that are congruent, we visually compared the individual gene genealogies for evidence of contradictory phylogenetic relationships. Although the incongruence length difference test (ILD; implemented as the partition length homogeneity test in PAUP) is still used in many phylogenetic studies to assess potential conflict among data sets, several properties of this test are known to generate misleading results, particularly when among-site rate variation is present (9,21,23,24,57). Because this is the case in our study, the ILD is probably not a reliable method to determine congruence of data sets stringently. We therefore relied on visual inspection to evaluate congruence between the individual gene genealogies. We found only three isolates with inconsistent placement between the gene trees: MA-6722 from *Poa annua* grouped with *C. cereale* clade A for the *SOD2* and combined trees, but with *C. cereale* clade B for the ITS and HMG; NJ-8467 from *Poa annua* grouped with *C. cereale* clade A for the ITS, HMG, and combined trees, but with *C. cereale* clade B for the *SOD2* tree; and NY-15182 from *Z. mays* grouped with *C. graminicola* for the HMG, *SOD2*, and combined trees, but with the *C. cereale* lineage for the ITS tree. Aside from these three individuals, the three gene genealogies all recovered monophyletic groups of *C. sublineolum*, *C. graminicola*, and *C. cereale* taxa, were topologically congruent with one another at these main clades, and were also in agreement with population subdivisions recovered through an analysis of three transposable element restriction fragment length polymorphism (RFLP) patterns (J. Crouch, B. B. Clarke, and B. I. Hillman, *unpublished data*). The individual data sets were therefore combined and used for multilocus analysis.

Analysis of the three gene sequences using the AIC in the computer program Modeltest determined that the evolutionary signature of each sequence was best modeled separately for the Bayesian analyses (Table 3). Although the genes ultimately reconstruct the same relationships between the ingroup taxa, a distinct set of parameters uniquely characterizes the evolutionary processes experienced by each gene. In particular, all three gene sequences showed different rates of variation among different nucleotide sites (among-site rate variation). To correct for the presence of among-site rate variation in the data set, which can cause the number of unobserved, multiple substitutions to be underestimated, the α shape parameter from the gamma distribution was incorporated into the models for each gene (which were established as priors in the Bayesian ML analysis) with α inversely proportional to the amount of among-site rate heterogeneity (when rates are equal, α = infinity). Among-site rate heterogeneity in the *SOD2* sequence was considerable (α = 0.8737); heterogeneity present in the other genes was also estimated to be high (ITS, α = 0.9962; HMG, α = 1.8284).

In the Bayesian likelihood analyses, one run plateaued at 60,000 generations, the second after 119,000 iterations, and the third reached stationarity at 53,000 (ln likelihood values: run 1: –8785.664 to –8699.69, average –8733.853; run 2: –8826.63 to –8740.83, average –8740.83; and run 3: –8794.062 to –8658.552, average –8697.4453). All three runs converged on the same topology, so all generations sampled from the posterior distribution were combined (83,539 trees) to produce a 75% consensus tree (Fig. 1) representing our hypothesis of descent for these taxa. The consensus tree had an ln likelihood score of –8,746.140; only clades with relatively strong support (>75% support for nodes from the posterior distribution) were retained.

Phylogenetic relationships. All phylogenetic analyses, separate and combined, using either the complete data set, the data set reduced to unique haplotypes, or the haplotype data set reduced to strictly compatible sites (83 total sites) resolved three distinct,

TABLE 2. Characteristics of the genomic regions used in this investigation

Gene	Exon regions	Intron regions	Nucleotide characters, excluding indels	Coded characters (indels)	Invariable nucleotide characters[a]	Variable, parsimony uninformative nucleotide characters[a]	Parsimony informative nucleotide characters[a]	MP tree length	Number of equally parsimonious trees
ITS	1	2	482	49	358	68	56 (12%)	357	44,034
HMG	2	1	211	13	170	14	26 (12%)	139	4
SOD	3	2	536	38	504	0	221 (41%)	518	72
Combined data set	6	5	1,229	130	1,032	82	303 (25%)	1,058	8,813

[a] Parsimony informative characterization of data set considers ingroup taxa only, outgroups are excluded.

TABLE 3. Best fit evolutionary models used in Bayesian analysis as determined by the Akaike information criterion (AIC) in Modeltest

ITS1/5.8S/ITS2 482 nucleotide characters	HMG (*MAT1-2*) 211 nucleotide characters	*SOD2* 536 nucleotide characters
SYM+G	HKY+G	HKY+G
A→C 1.0161	Frequency A = 0.305	Frequency A = 0.2352
A→G 1.2515	Frequency C = 0.269	Frequency C = 0.2575
A→T 1.0096	Frequency G = 0.224	Frequency G = 0.3088
C→G 1.7520	Frequency T = 0.201	Frequency T = 0.1986
C→T 2.7189	....	...
G→T 1.000	...	...
...	Ti/Tv ratio = 1.2350	Ti/Tv ratio = 1.5889
Invariable = 0	Invariable = 0	Invariable = 0
Gamma	Gamma	Gamma
α = 0.9962	α = 1.8284	α = 0.8737

Appendix 2: Publication version of Chapter 1

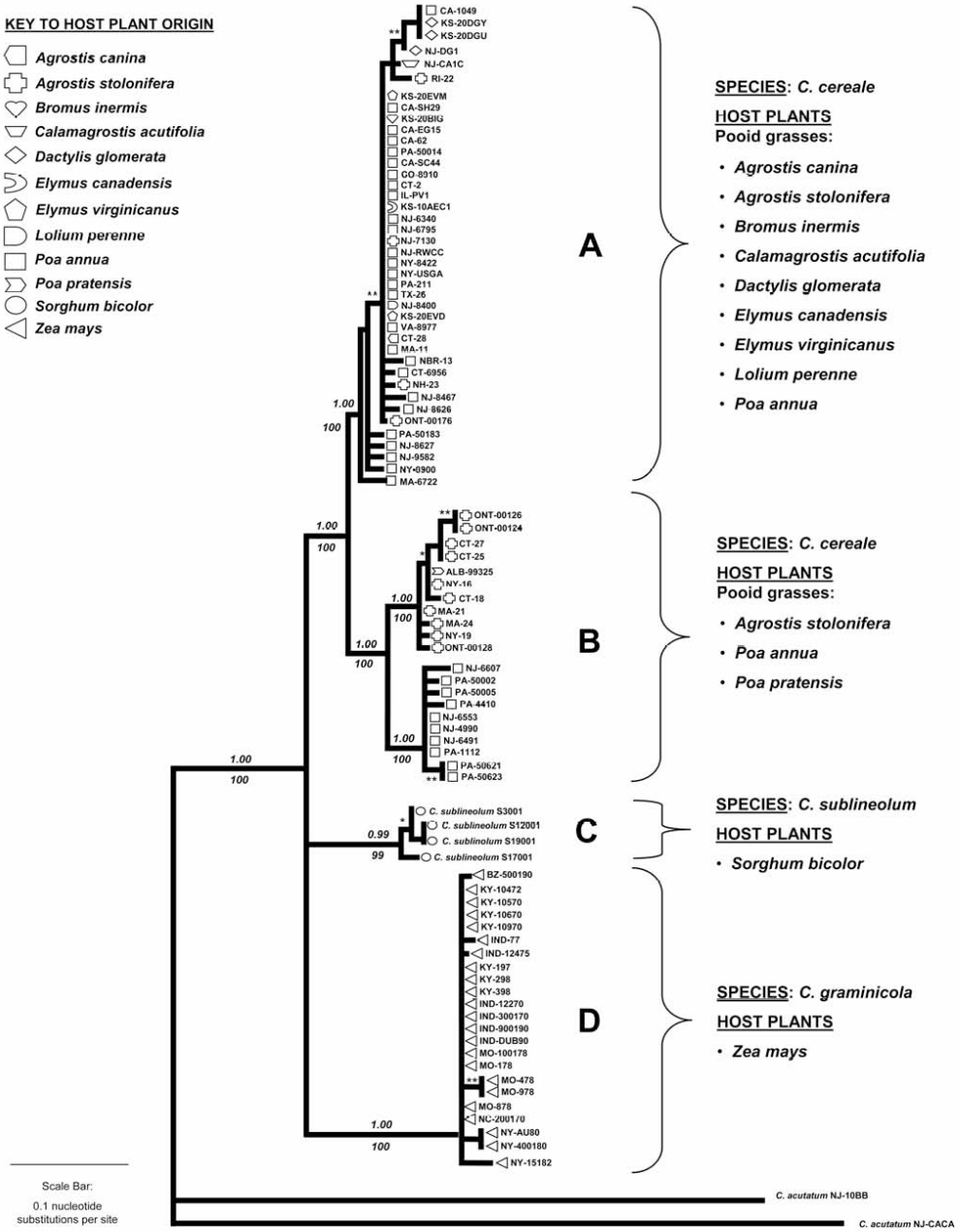


Fig. 1. Seventy-five percent majority rule Bayesian maximum likelihood phylogenetic tree constructed from the 83,539 trees sampled from the posterior distribution representing the well-supported hypothesis of descent recovered in the multilocus sequence analysis. The scale bar indicates divergence, measured in nucleotide changes per site. All groups illustrated by this tree are supported by posterior probabilities of at least 0.75. Posterior probabilities supporting the topology are shown above the nodes where space permits; elsewhere on the tree two asterisks were supported by posterior probabilities of >99%; branches with one asterisk were supported by posterior probabilities of >95%. Maximum parsimony analysis also recovered this same basic phylogeny; bootstrap values from 1,000 pseudoreplicates are shown below the nodes. Symbols next to taxa indicate host plant origin; the key to host plant origin is to the left of the phylogeny.

Appendix 2: Publication version of Chapter 1

well-supported groups: (i) a monophyletic group comprising isolates from Pooideae grasses (clades A and B, Fig. 1), (ii) a group comprising isolates from sorghum (*C. sublineolum*, clade C), and (iii) a group comprising isolates from maize (clade D). In the *SOD2* and the multilocus tree, a further subdivision was observed: the *C. cereale* clade subdivided into unique lineages (clades A and B). *C. cereale* clade B further diverged into two smaller groupings. These groups were supported by both bootstrap values and posterior probabilities greater than 95%. The phylogenetic trees recovered under both parsimony and Bayesian likelihood optimality criteria are in agreement at all of the main clades. Figure 1 shows the 75% consensus tree constructed from the 83,539 trees sampled from the Bayesian posterior distribution; both the posterior probabilities and bootstrap support values are plotted at the nodes (the MP tree is not illustrated separately due to topological congruence with the ML phylogeny; groups not supported by at least a value of 0.75 by both bootstrap and posterior probabilities were collapsed). The results of the molecular phylogenetic analysis reconstructs a species history characterized by the presence of morphologically similar sister species: *C. graminicola*, which is pathogenic to corn, and a group of *C. cereale* lineages found on a wide range of pooid grasses, including cultivated C3 turfgrasses.

A much less well-supported association between *C. graminicola* and *C. sublineolum* was consistently recovered in all of the phylogenetic analyses, with these two species always forming a monophyletic group, separate from the *C. cereale* group. This association between the cereal-derived lineages was not, however, entirely well supported, with a posterior probability of only 0.63 in the Bayesian consensus tree. Analysis of the haplotype data set using the split decomposition network reconstruction method (Fig. 2) also estimates a close relationship between *C. graminicola* and *C. sublineolum*, but confidence in a recent common ancestry for these taxa was extremely low (bootstrap = 56).

Reticulation and incompatibility in the data set. A total of 57 unique multilocus haplotypes were identified from the ingroup sample, with 611 nucleotides distinctively establishing the haplotypes after the removal of introns and any positions violating the infinite sites model. The split-decomposition splits-graph used to visualize the presence of reticulate evolution in the haplotype data set (Figs. 2 and 3) recovered tree-like relationships between the main species clades also estimated by the phylogenetic investigation: *C. sublineolum* from sorghum, *C. graminicola* from corn, and the *C. cereale* species group from the subfamily Pooideae grasses. Within the clades, on the intraspecific level, a reticulating network of relationships clearly emerged: a single loop was present among the *C. graminicola* haplotypes (five haplotypes, 23 individuals), a single loop connected the *C. sublineolum* haplotypes (three haplotypes, four individuals), and more complex patterns were found individually among the *C. cereale* lineages (Fig. 3). The two main *C. cereale* phylogenetic clades A and B were inferred in the splits-graph, with reticulate lineages occurring only below the level of these groups. *C. cereale* clade A consisted of 27 unique haplotypes from 59 individuals, all radiating from haplotype 5 (H5), which, given its central position in the splits-graph, is likely the ancestral haplotype for the *C. cereale* lineages. This interpretation is further supported by the observation that H5 contained the largest number of *C. cereale* isolates (22), and was widely distributed throughout the entire range of our sample. The majority of the haplotypes were tightly interrelated in clade A,

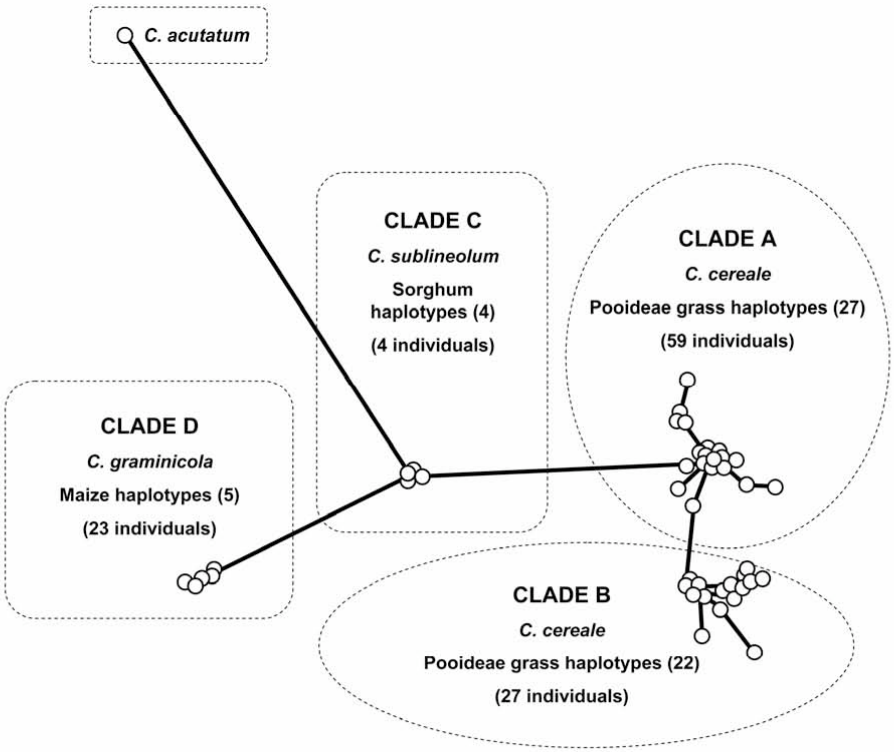


Fig. 2. Evolutionary relationships between the sampled taxa, inferred using split-decomposition network analysis as described in the text.

Appendix 2: Publication version of Chapter 1

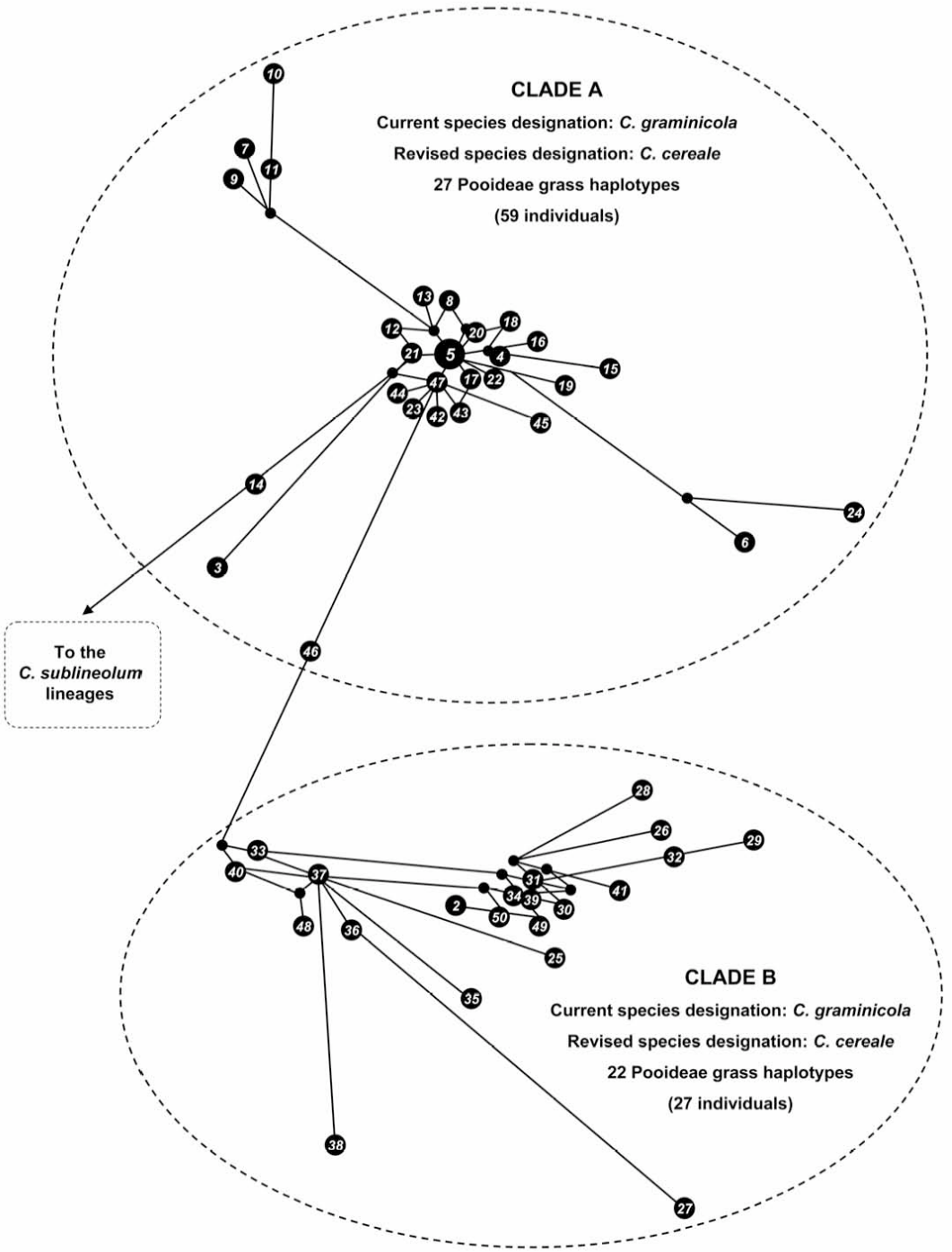


Fig. 3. An enlarged view of the splits-graph topology (Fig. 2), illustrating *Colletotrichum cereale* clades A and B. The small filled circles represent the positions o inferred intermediate haplotypes.

Appendix 2: Publication version of Chapter 1

with only a few divergent lineages. Based upon its position within the splits-graph, clade A haplotype 47 also appeared to be a central lineage among the group, and served as the connecting point between *C. cereale* clades A and B. Haplotype 46, represented by *C. cereale* isolate NJ-8467 from *Poa annua*, was positioned directly between clades A and B, which is consistent with the observed conflict between gene genealogies for this isolate (ITS, HMG, and multilocus = clade A; *SOD2* = clade B). Collectively, these data suggest that this haplotype is a hybrid between the two *C. cereale* lineages. Haplotype 3, consisting of *C. cereale* isolate MA-6722, while similarly characterized by conflicting gene trees in the phylogenetic analyses (*SOD2* and multilocus tree = clade A; ITS and HMG = clade B), appears in the splits-graph to be a highly divergent taxon, but uniquely a member of the clade A lineage.

The haplotypes of *C. cereale* clade B, although numerically less prevalent in our sample, were nevertheless a much more genetically diverse assemblage, with 22 haplotypes resulting from 27 isolates interwoven into a complex pattern of cycles. While the phylogenetic analyses support the split of clade B into two distinct lineages, the splits-graph was unable to recover the subdivision, suggesting that these taxa are all of the same species.

Analysis of the haplotype data set by means of the compatibility matrix and Hudson's four gamete test identified the majority of characters as being incompatible; these characters were subsequently removed from the data set. The resultant 83 character data set was used to further investigate species-level boundaries through MP phylogenetic analysis. This reduced MP phylogeny (data not shown) confirmed the splits-graph topology: *C. sublineolum* from sorghum, *C. graminicola* from corn, and *C. cereale* from the Pooideae grasses are each individual species; however, the species level divergence of *C. cereale* clades A and B found in all other analyses was not recovered.

Taxonomy. Based upon the unique patterns of fixed nucleotide differences in DNA sequence data at the *ITS1/5.8S/ITS2*, *MAT1-2*, and *SOD2* nuclear loci, it is evident that *Colletotrichum* isolated from the pooid grasses examined in this study form a species taxon distinct from *C. graminicola*, which is limited to the fungus pathogenic to maize host plants. Based upon these nucleotide data, we formally propose to resurrect and emend the species *C. cereale* and to emend the species descriptions of *C. graminicola* and *C. sublineolum*.

Colletotrichum cereale Manns, Ohio Agric. Exp. Stn. Bull. 203:207 (1909).
Teleomorph. Unknown.

Emended description. Inhabits grasses of the subfamily Pooideae; pathogenic and sometimes particularly aggressive on *Poa annua*, *Poa pratensis*, *P. sapina*, *Agrostis stolonifera*, *Agrostis canina*, and *Lolium perenne*. *C. cereale* is uniquely described by the following fixed nucleotide characters (positions are relative to the nucleotide sequence alignment of the epitype specimens): *ITS1/5.8S/ITS2* positions 15 (C), 39 (T), 56 (C), 98 (C), 104 (C), 117 (G), 145 (C), 165 (A), 178 (T), 409 (A), 517 (T), 522 (C); *MAT1-2* positions 2 (A), 54 (C), 79 (A/T), 87 (G), 97 (A), 101 (C), 109 (A), 123 (T), 129 (G), 131 (A), 154 (C), 163 (C), 172 (A), 199 (A); and *SOD2* positions 35 (C/T), 45 (A), 47 (C), 53 (A), 58 (A), 61 (G), 70 (C), 73 (A), 77 (T), 88 (C), 119 (G), 127 (C), 133 (C), 136 (G), 142 (C), 151 (C), 169 (A), 175 (C), 205 (C), 211 (C), 226 (G), 244 (A), 247 (C), 259 (C), 262 (C), 268 (C), 280 (T), 292 (G), 332 (G), 355 (C), 373 (G/A), 376 (C), 382 (C), 386 (G), 394 (T), 403 (C), 451 (C), 454 (C), 455 (G), 500 (T), 501 (G), 512 (C), 518 (T), 519 (T), 520 (G), 521 (G), 522 (A), 525 (G), 534 (A), 537 (C), 541 (C), 545 (T), 546 (T), 547 (C), 551 (C), 558 (T), and 571 (G).

Colonies on PDA under constant illumination highly variable; usually form dark mat of tight setae masses across agar surface; commonly heavy conidia accumulation gives orange cast to brown/black culture. Some cultures exhibit thin layer of hyphae along agar surface, copious amounts of conidia give orange appearance. Some cultures possess fluffy aerial mycelium growing over setae, producing gray appearance; generally as colonies age mycelia overtakes entire culture surface. Hyphae septate, normally hyaline, sometimes dark brown when present at the base of setae, 1.0 to 6.5 µm, often guttulate. Conidia falcate or fusiform, apices acute, individually hyaline but appear salmon-orange en mass, may be mono-, bi-, or up to seven-guttulate or oil drops may be absent from the cytoplasm, measuring 6.0 to 33.8 µm × 2.2 to 6.3 µm with an average of 23.3 µm × 3.4 µm. Germinating conidia

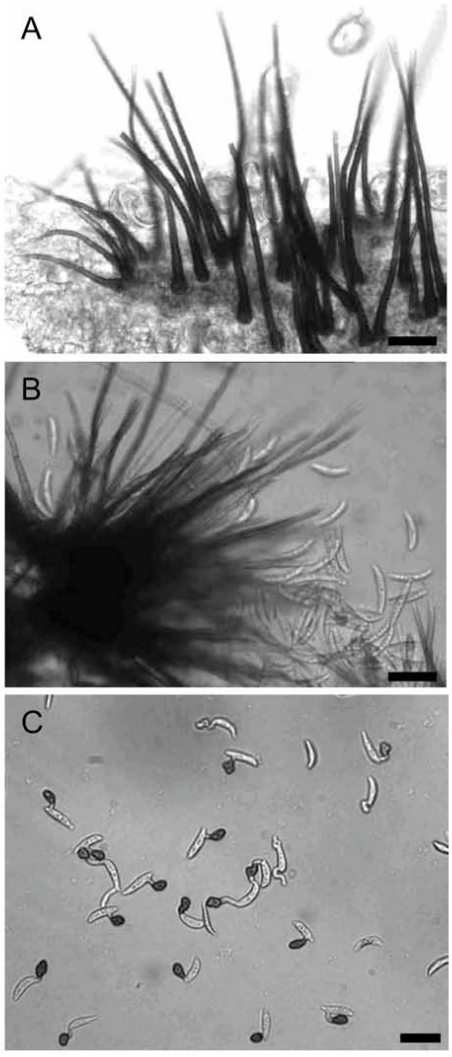


Fig. 4. *Colletotrichum cereale* morphological structures (Bar = 25 µm). **A,** Heavily melanized setae emerging from an acervulus on the leaf of *Bromus secalinus* from the lectotype specimen (NYBG-305598); **B,** setae and conidia, grown on potato dextrose agar from epitype strain KS-20BIG; and **C,** germinating conidia, with the formation of germ tubes generally (but not always) leading to the formation of heavily melanized appressoria.

Appendix 2: Publication version of Chapter 1

form single or infrequently two hyaline germ tubes that terminate in dark brown/black appressoria, sometimes appressoria absent; germ tubes separated from appressoria by septa (Fig. 4), occasionally appressoria forms directly from conidia. Appressoria rounded and smooth or irregular or lobate or multi-lobate, measuring 8.5 to 11.6 µm × 6.5 to 10.2 µm. Setae develop from dark brown, tight masses of hyphae (Fig. 4); in culture dense, rounded masses of setae form, (Fig. 4) sometimes partly or completely covered in thick deposition of conidia. Individual brown-black setae separated from hyphae by septa. Setae base swollen or not swollen, irregularly septate with up to 7 septa, measuring 32 to 120 µm × 6 to 8 µm at base, tapering at tip.

Type specimen. Examination of herbarium syntype specimens confirmed the presence of acervuli on the grass hosts, in association with black-brown setae that are characteristic of *Colletotrichum*. Very few conidia were observed; those that were present were somewhat shriveled and without cytoplasmic contents, falcate in shape, measuring 20 to 25 µm × 3 µm. Morphological comparison of the syntypes of *C. cereale* and the published descriptions and illustrations against our recent samples obtained from pooid grasses confirmed that the *Colletotrichum* isolated from pooid grasses in the present study are within the range of morphological variation consistent with the *C. cereale* species description. Although we attempted to extract DNA from the herbarium material, we were not successful, as the fungal tissue was in close association with the plant tissue, and relatively little fungal tissue was present. Attempts to revive the specimens on PDA were also unsuccessful; therefore comparisons of contemporary *Colletotrichum* strains with syntype specimens are necessarily confined to morphological estimations. Four of the New York Botanical Garden specimens were catalogued in the herbarium database as type specimens (although the published description never makes such a specification): 305598 on *Bromus secalinus*, 305599 on *Avena sativa*, 305595 on *Arrhenatherum elatius*, 305596 on *Secale cereale*, and 305597 on *Triticum vulgare*. A holotype was not designated; we therefore establish 305598 (*Bromus secalinus*) as the lectotype for the species, and 305599, 305595, 305596, and 305597 are specified as paratypes. To facilitate species interpretation using the molecular characters described in this study, we designate KS-20BIG, NJ-6795, PA-5062-3, and NJ-4990 as epitypes; cultures of the epitype strains have been deposited in the American Type Culture Collection (ATCC), Manassas, VA, and the Centraalbureau voor Schimmelcultures (CVS), Utrecht, the Netherlands.

Comments. *C. cereale* also possesses several fixed autapomorphic characters that serve to uniquely distinguish the species from *C. graminicola* (positions are relative to the nucleotide sequence alignment of the epitype specimens): *ITS1/5.8S/ITS2* positions 54 (no base), 69-72 (no bases), 100 (C), 111 (A), 115 (G), 121 (no base), 158 (no base), 404-406 (no bases), 525 (no bases); *MAT1-2* positions 115 (no base), 120 (G); and *SOD2* positions 18-19 (AA), 25-27 (no bases), 29 (T), 56 (A), 60 (no base), 78 (no base), 82 (no base), 85 (T), 503 (no base), 508 (T), 562 (A), and 567 (no base).

In general, *C. cereale* is morphologically very similar to *C. graminicola*. Two continuous morphological characters (conidia length and hyphopodium area) have recently been suggested to differentiate approximately between *C. graminicola* and *C. cereale* (12,25); however, the variability and overlap of range in these types of quantitative characters precludes diagnosis of these taxa uniquely and consistently on this basis (61). Conidia in these two species are sometimes different in length, with the conidia of *C. graminicola* typically, but not always, larger than those from *C. cereale* (6,12,25,54, our data); overall, *C. cereale* conidia grown on PDA measured 6.0 to 33.8 µm, while *C. graminicola* conidia ranged from 13.0 to 44.0 µm. Hyphopodium (mycelial appressoria) area is also notably different in the two species (12,25), with the hyphopodium area of *C. graminicola*, on average, significantly larger than those of *C. cereale*, but this continuous character also is present in overlapping ranges (*C. cereale* 63.8 to 315.6 µm^2 and *C. graminicola* 136.7 to 1,027 µm^2) (12,25,61).

Colletotrichum graminicola (Ces.) G.W. Wilson, Phytopathology 4:110 (1914) (as "*graminicolum*").
- = *Dicladium graminicola* Ces., Flora 35:398 (1852) (as "*graminicolum*").
- = *Vermicularia graminicola* (Ces.) Westd., Bull. Acad. Roy. Brux. 12: n. 7 (1861).
- = *Steirochaete graminicola* (Ces.) Sacc. Syll. Fung. 4:316 (1886).
- = *Colletotrichum zeae* Lobik, Trudy severo-kavkazskogo Instituta Zashchity Rastenii 1(2):39 (1933).
- = *Colletotrichopsis graminicola* (Ces.) Muntaola, Rev. Argent. Agron. 19:220 (1952).
- = *Colletotrichum graminicola* f. sp. *zeae* Messaien, Lafon & Malot, Ann. Epiphyt., ser. C., 10:454 (1959).

Teleomorph. ***Glomerella graminicola*** Politis, Mycologia 67:56-72 (1975).

Emended description. Parasitic on *Zea mays*. *C. graminicola* is uniquely distinguished by the following synapomorphic fixed nucleotide characters (positions are relative to the nucleotide sequence alignment of the epitype specimens): *ITS1/5.8S/ITS2* positions 15 (T), 39 (C), 56 (T), 92 (G), 98 (C), 99 (A), 104 (T), 117 (C), 145 (T), 165 (G), 178 (C), 409 (T), 517 (C), 522 (T); *MAT1-2* positions 2 (T), 54 (T), 79 (G), 87 (A), 97 (G), 101 (T), 109 (G), 123 (C), 129 (A), 131 (C), 154 (T), 163 (T), 172 (G), 199 (G); and *SOD2* positions 35 (A), 45 (G), 53 (T), 58 (T), 61 (A), 70 (G), 73 (G), 77 (C), 88 (T), 119 (A), 127 (T), 133 (T), 136 (A), 142 (G), 151 (T), 169 (G), 175 (A), 205 (T), 211 (A), 226 (T), 244 (T), 247 (T), 259 (T), 262 (G), 268 (A), 292 (C), 332 (A), 255 (T), 373 (T), 376 (T), 382 (T), 386 (A), 394 (C), 403 (T), 451 (T), 454 (T), 455 (A), 500 (C), 501 (A), 512 (T), 518 (A), 519 (A), 520 (A), 521 (A), 522 (G), 525 (A), 534 (G), 537 (A), 541 (T), 545 (A), 546 (A), 547 (T), 551 (T), 558 (C), and 571 (A).

Epitype establishment. The lectotype established for this species was examined by Sutton (58) from *Z. mays* (IMI83255). To facilitate species interpretation using the molecular characters described in this study, we designate MO-100178 from *Z. mays* (also known as *C. graminicola* isolate M1.001) as an epitype; a culture of this specimen has been deposited in the ATCC and the CVS.

Comments. The reader is referred to any of several excellent morphological studies of *C. graminicola* sensu stricto Sutton that have been published (10,25,58–62). In general, *C. graminicola* is morphologically quite similar to *C. cereale* except for the two continuous quantitative conidial and hyphopodial characters noted above, but the overlapping range of these morphological characters precludes their usage for consistent and unique diagnosis of these species.

C. graminicola also possesses several fixed autapomorphic characters that can serve to uniquely distinguish the species from *C. cereale* (positions are relative to the nucleotide sequence alignment of the epitype specimens): *ITS1/5.8S/ITS2* positions 54 (T), 69-72 (TCCG), 100 (no base), 111 (no base), 115 (no base), 121 (G), 158 (A), 404-406 (GTA), 525 (C); *MAT1-2* positions 115 (C), 120 (no base); and *SOD2* positions 18-19 (no bases), 25-27 (AAC), 29 (no base), 47 (no base), 56 (no base), 60 (T), 78 (C), 82 (C), 85 (no base), 500 (C), 508 (no base), 562 (no base), and 567 (C).

Colletotrichum sublineolum Henn. *apud* Kabat & Bub. Fungi imp. exs. 186 (1905) (as "*sublineola*").
- ≡ *Colletotrichum graminicola* f. sp. *sorghi* Messaien, Lafon & Malot, Ann. Epiphyt., ser C., 10:456 (1959).
- = *Colletotrichum graminicola* var. *zonatum* Rajasab & Ramal., Curr. Sci. 50(1):34 (1981).

Teleomorph. Unknown.

Appendix 2: Publication version of Chapter 1

Emended description. Parasitic on *Sorghum bicolor* and *Sorghum halapense*. *C. sublineolum* is uniquely distinguished by the following synapomorphic fixed nucleotide characters (positions are relative to the nucleotide sequence alignment of the epitype specimens): *ITS1/5.8S/ITS2* positions 39 (T), 67 (T), 93 (C), 98 (T), 99 (C), 105 (C), 112 (C), 131 (G) 132 (A), 133 (G), 135 (A), 136 (T), 137 (A), 139 (G), 142 (A), 143 (A), 147 (T), 178 (T), 417 (T), 419 (A), 420 (C); *MAT1-2* positions 2 (A), 54 (C), 66 (T), 77 (A) 79 (A) 87 (G), 88 (A), 95 (G), 106 (G), 130 (A), 145 (A), 154 (C), 163 (C), 181 (T), 202 (T); and *SOD2* positions 40 (C), 42 (T), 44 (T), 48 (T), 88 (C), 115 (A), 125 (G), 142 (C), 148 (A), 151 (C), 166 (T), 175 (C), 184 (T), 187 (A), 205 (C), 211 (C), 226 (G), 244 (A), 268 (C), 332 (G), 355 (C), 367 (A), 373 (G), 376 (C), 382 (C), 386 (C), 403 (C), 408 (T), 455 (G), 457 (C), 460 (A), 495 (A), 498 (G), 504 (G), 505 (G), 509 (T), 512 (G), 518 (G), 519 (G), 524 (A), 530 (G), 534 (A), 537 (T), 538 (C), 541 (G), and 563 (A).

Epitype establishment. To facilitate species interpretation using the molecular characters described in this study, we designate S3.001 from *Sorghum bicolor* as an epitype; a culture of this specimen has been deposited in the ATCC and the CVS.

Comments. *C. sublineolum* also possesses several fixed autapomorphic characters that serve to uniquely distinguish the species from *C. graminicola* (positions are relative to the nucleotide sequence alignment of the epitype specimens): *ITS1/5.8S/ITS2* positions 25 (T), 69-72 (no bases), 100 (C), 101-102 (TC), 107 (no bases), 111 (A), 113-115 (GGG), 119-120 (CG), 138 (A), 414-416 (no bases), 514 (T), 522 (no base); *MAT1-2* positions 105 (C), 116 (no base), 120 (G); and *SOD2* positions 35 (no base), 37 (no base), 39 (A), 47 (T), 50 (no base), 52 (no base), 503 (no base), 508 (G), and 544-546 (GGA).

We have not listed a teleomorph state for the species; in a review of the species taxonomy, Sutton suggests that the teleomorph may be *Glomerella cingulata* var. *sorghicola* Saccas (62).

It is worth emphasizing that although *C. graminicola* is still frequently employed to describe *Colletotrichum* from sorghum, Sutton's 1980 emendment of the species description rendered such application of the name illegitimate. *C. sublineolum* is the validly established taxon describing *Colletotrichum* associated with sorghum; this distinction has been repeatedly confirmed through morphological examinations (10,12,25,59–61), host range and pathogenicity testing (6,12,20,36,40), and molecular analyses (12,19,25,31,45,52,55,56,67,68, this study).

DISCUSSION

The purpose of this study was to explore the evolutionary history of the *C. graminicola* lineages that cause anthracnose disease in turfgrasses, in particular, we sought to identify the point at which interspecific boundaries have been erected between taxa, defining the extreme limits of gene flow and population-level processes.

Species boundaries. The phylogenetic analyses performed in this study generated an extremely well-supported hypothesis of evolutionary descent for the taxa currently recognized as *C. graminicola* sensu lato G.W. Wilson and conclusively established that there are two species within this circumscription. Although we employed several methods to identify reticulate evolution among the taxa, the evidence is overwhelmingly in favor of a long-standing biological isolation between these two distinct, monophyletic groups.

In light of the pattern of fixed nucleotide differences and the unique host plant associations observed between lineages of *Colletotrichum* from corn and from those isolates obtained from the C3 grasses in the subfamily Pooideae, the species level phylogenetic divergence present in *C. graminicola* sensu lato G.W. Wilson is unambiguous. We therefore propose to resurrect the species *C. cereale* to describe *Colletotrichum* from pooid grasses, and *C. graminicola* (Ces.) G.W. Wilson sensu stricto Sutton 1980 is reserved for the fungus pathogenic to corn, with each species uniquely characterized by their molecular identities at the *ITS1/5.8S/ITS2*, *MAT1-2*, and *SOD2* loci and their host range. These physiological host range limitations are well documented in the literature (2,6,12,20,36,38,40,41,43,44,72,73) and correspond with the emended phylogenetic species described in this paper. Since *C. cereale* likely has a much more extensive host range as a pathogen than that which is explicitly described in this manuscript, we emphasize that it is the pattern of fixed differences on the molecular level and the association with grasses of the Pooideae that serve uniquely to distinguish this species, rather than strict pathogenicity criteria. This is an important distinction: *C. cereale* transcends its ability to induce anthracnose disease symptoms in a given host plant, since the species is capable of inhabiting many pooid grasses without inducing disease in the host.

Even in our most conservative analyses, where only the most reliable, unconflicted nucleotide characters are considered, these species fully meet the criteria of the genealogical concordance phylogenetic species concept (64), which is capable of efficiently and objectively marking species boundaries in asexually reproducing organisms such as *Colletotrichum* that defy characterization based upon morphological or reproductive criteria. The specific designation of these groups is dictated by taxonomic priority: *C. graminicola*, as *Dicladium graminicolum*, was first described by Cesati (15) in 1852 from the stems of corn and barnyard grass (*Echinochloa crus-galli*); the first formal description of *Colletotrichum* from a pooid grass did not occur until 1909 (54) when Selby and Manns first described *C. cereale*. Most of the taxonomic uncertainty surrounding the circumscription of *C. graminicola* completely by-passed members of the genus found in association with pooid grasses and instead concentrated on those strains responsible for economically important levels of disease in the host plant: corn, sorghum, and sugarcane. Even Sutton, the recognized authority in *Colletotrichum* systematics, has evaluated only two pooid strains as *C. graminicola* (from *Avena sativa*) using conidial measurements. Herbaria material inspected from *Bromus* spp., *Calamagrostis epigeoios*, *C. villosa*, *C. neglecta*, *Lolium perenne*, and *Poa annua* were accepted by Sutton to be *C. dematium* since a distinction between these taxa could not be made based on morphological structures examined (58). The study of Selby and Manns that served to erect *C. cereale* was not cited or discussed in any of Sutton's publications (58–62).

Because our current research goals are focused on the population dynamics of *Colletotrichum* from pooid grass hosts rather than conducting an expansive taxonomic survey of the genus, we have not sampled *Colletotrichum* from the other C4 grasses evaluated by Wilson (*Panicum* spp., *Echinochloa crus-galli*), leaving the species status of these members of the genus *Colletotrichum* still undefined; Sutton's 1980 treatment of *C. graminicola* precludes the legitimate use of the name to describe these taxa. In a review of the taxonomy of the genus *Colletotrichum* in 1992 (62), Sutton introduced the idea that a still unnamed species closely related to *C. falcatum* occurs on grass hosts other than sugarcane, basing his suggestion upon earlier morphological studies of samples from several "miscellaneous" C4 grass genera of the Panicoideae including *Andropogon*, *Digitaria*, *Echinochloa*, *Eleusine*, *Eragrostis*, *Miscanthus*, *Panicum*, and *Rottboellia* (58). Molecular data presented in two recent phylogenetic analyses (12,25) corroborate Sutton's hypothesis, providing additional evidence that a new species will need to be erected to describe properly the *Colletotrichum* associated with *Echinochloa* spp., since these taxa do not conform phylogenetically to any of the species now described from graminaceous host plants.

One particular question that remains unresolved is the relationship between *C. graminicola* and *C. sublineolum*: have they recently diverged from a common ancestor? This is an intriguing

Appendix 2: Publication version of Chapter 1

premise that continually arises in all of the evolutionary inferences, albeit with minimal support. For several years these two species, along with *C. falcatum* and *C. caudatum*, were considered to be varietal forms of a single species (5,43,44), and the unique status of *C. sublineolum* is still rejected by many researchers of sorghum anthracnose (e.g., 69). Is this association between species real, or is it an artifact of ancestral polymorphisms? To fully consider this possibility, a substantially larger portion of the genome should be sampled, with an extended sampling from these taxa along with their sister species from other C4 grass hosts.

Unresolved, potentially species-level divergences. We envision that *C. cereale* sensu lato Crouch, Clarke, and Hillman may eventually need to be further subdivided—minimally into two species, and potentially into as many as four distinct species based solely upon this limited sampling. We regard the description of *C. cereale* to be a species group rather than a single homogeneous species, but the data from the present study are inadequate for the purpose of formally proposing subdivision of the group. Both ML and MP phylogenetic inferences and the splits-graph strongly support the divergence of *C. cereale* clades A and B, but the phylogeny of the data set reduced of all potentially incompatible sites does not recover these lineages as independent entities. While this discrepancy is likely due to the elimination of a great many of the informative characters from the analysis rather than continued gene flow, we cannot at this time confidently suggest the establishment of two unique species based upon these data alone, since the eliminated characters may have yielded erroneous species phylogenies. It is not surprising that a great deal of potentially misleading data can be found when considering a group of organisms so close to the population level, as any number of processes can result in inconsistent, misleading conclusions due to incomplete lineage sorting, recombination, and hybridization. We expect that the application of several analytical techniques—vegetative compatibility, host range analysis, pathogenicity, and virulence—will be required before these unique groups can be defined in a substantive, biologically meaningful manner. This work merely represents the first step in an ongoing process; nevertheless, it does provide a vital framework for future experimental work.

The evolutionary history of *C. cereale*. In any phylogenetic analysis, the assumption that only a single phylogeny underlies the evolution of the population sampled is violated by the presence of recombination. Recombination is a truly creative force in organismal biology, but has the unfortunate side effect of completely confounding the accurate estimation of phylogenies when present (14,48,49 for reviews). This is particularly true in sexual species and, as our data will attest, in putatively asexual entities like those found in the genus *Colletotrichum*. It is evident that although *C. cereale* may for the most part propagate in an asexual fashion, recombination between taxa has occurred, as independently estimated by the split-decomposition network and compatibility matrix analysis; results from transposon RFLP fingerprinting assays also confirm the existence of mosaic genotypes in the species group (J. Crouch, B. B. Clarke, and B. I. Hillman, *unpublished data*). Although the three individual gene genealogies considered in this study were topologically congruent, suggesting a clonal, nonrecombining species history, overall our data provide evidence for the presence of recombination that is not reflected in the tree topologies. Many experts in the field of evolutionary biology recommend using a combination of assays to detect recombination in order to maximize the possibility of identifying recombination with a minimum of false positives (50 for review). In our research, three separate analyses detected the potential for recombination between *C. cereale* clades A and B; we believe these data are sufficient to recommend further analysis before establishing these clades as phylogenetic species when they may in fact only represent divergent populations.

Regardless of whether one accepts the clades of *C. cereale* as individual species or simply highly divergent populations, there are undeniably two unique lineages of this fungus causing disease in turfgrasses maintained as golf course greens. The clade A isolates are geographically widespread and were isolated from numerous turfgrass and noncultivated C3 pooid grass species; they are the numerically dominant form of *C. cereale* in the sample discussed in this manuscript and in our entire culture collection. To date, over 200 additional specimens of *C. cereale* in our culture collection have been identified as part of clade A through the application of a PCR-based screening protocol using lineage-specific transposable element primers, in some instances supported by sequence analysis. Also included in *C. cereale* clade A are several isolates from wheat plants from the Midwest United States, as inferred from *SOD2* and *ITS* sequences (J. Crouch, B. B. Clarke, and B. I. Hillman, *unpublished data*). In contrast, the clade B lineage is quite rarely encountered and appears to have a much more restricted host range and geographic distribution. Additionally, there are several hints of an association between host plant derivation and the *C. cereale* clade B lineage: the phylogenetic tree topology in particular shows an almost perfect division between *Agrostis stolonifera* and *Poa annua* isolates within this lineage. Whether the peculiarities of clade B with respect to distribution and host range are based on real phenomena or have arisen due to sampling bias will require further study.

One of the most important conclusions generated by these analyses is the identification of the very common and widely distributed H5 ancestral haplotype in the *C. cereale* clade A population. The proposition that all North American *C. cereale* lineages are ultimately derived from the H5 haplotype, regardless of their host-plant derivation, could have very important implications for the development of effective disease control strategies.

ACKNOWLEDGMENTS

We are grateful to Lisa Vaillancourt for her suggestion to use the *C. graminicola* mating-type locus as a phylogenetic marker in our analysis and for sharing a pre-publication version of reference 25 while this manuscript was in review. We thank Richard Buckley and Sabrina Tirpak of the Rutgers Disease Diagnostic Clinic for identifying and supplying *Colletotrichum*-infected turfgrass specimens, and to several individuals who generously provided fungal cultures for use in this study: Jim Baird, Gary Bergstrom, Tom Hsiang, Noel Jackson, Randy Kane, Peter Oudemans, Wakar Uddin, Lisa Vaillancourt, and Frank Wong. Special thanks to the herbarium staff of the New York Botanical Garden for their assistance with herbarium specimens. We also thank Karen Garrett and Ned Tisserat for the use of laboratory facilities and assistance with sampling conducted at the Konza Prairie Biological Station (a preserve of The Nature Conservancy managed by the Division of Biology, Kansas State University), and Karl Kjer for assistance with the data sets, phylogenetic analyses, and for comments on the manuscript. We also thank Tom Harrington, Jim White, and three anonymous reviewers, who provided valuable discussions, comments, and criticisms that greatly improved the quality of this manuscript. J. A. Crouch was supported by a Rutgers Excellence Fellowship and a Land Institute Natural Systems Agriculture Fellowship. This project was funded by the Rutgers Center for Turfgrass Science.

LITERATURE CITED

1. Akaike, H. 1974. A new look at the statistical model identification. IEEE Trans. Automatic Control 19:716-723.
2. Ali, M. M. 1962. Factors influencing pathogenicity of three isolates of *Colletotrichum graminicola* on wheat. Mycopath. Mycol. Appl. XIX:161-166.
3. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. J. Mol. Biol. 215:403-410.
4. Arie, T., Christiansen, S. K., Yoder, O. C., and Turgeon, B. G. 1997. Efficient cloning of ascomycete mating genes by PCR amplification of the conserved *MAT* HMG box. Fungal Genet. Biol. 21:118-130.
5. Arx, J. A. v. 1957. Die arten der gattung *Colletotrichum* Cda. Phytopathol. Z. 29:413-468.

Appendix 2: Publication version of Chapter 1

6. Backman, P. A., Landershoot, P. J., and Huff, D. R. 1999. Variation in pathogenicity, morphology, and RAPD marker profiles in *Colletotrichum graminicola* from turfgrasses. Crop Sci. 39:1129-1135.
7. Balardin, R. S., Smith, J. J., and Kelly, J. D. 1999. Ribosomal DNA polymorphism in *Colletotrichum lindemuthianum*. Mycol. Res. 103:841-848.
8. Bandelt, H. J., and Dress, A. W. 1992. Split decomposition: A new and useful approach to phylogenetic analyses of distance data. Mol. Phylogenet. Evol. 1:242-252.
9. Barker, F. K., and Lutzoni, F. M. 2002. The utility of the incongruence length difference test. Syst. Biol. 51:625-637.
10. Baxter, A. P., van der Westhuizen, G. C. A., and Eicker, A. 1983. Morphology and taxonomy of South African isolates of *Colletotrichum*. South African J. Bot. 2:259-270.
11. Brasier, C. M. 2001. Rapid evolution of introduced plant pathogens via interspecific hybridization. BioScience 51:123-133.
12. Browning, M., Rowley, L. V., Zeng, P., Chandlee, J. M., and Jackson, N. 1999. Morphological, pathogenic, and genetic comparisons of *Colletotrichum graminicola* isolates from *Poaceae*. Plant Dis. 83:286-292.
13. Cannon, P. F., Bridge, P. D., and Monte, E. 2000. Linking the past, present, and future of *Colletotrichum* systematics. Pages 1-20 in: *Colletotrichum*: Host Specificity, Pathology, and Host–Pathogen Interaction. D. Prusky, S. Freeman, and M. B. Dickman, eds. The American Phytopathological Society, St. Paul, MN.
14. Carbone, I., and Kohn, L. M. 2004. Inferring process from pattern in fungal population genetics. Pages 1-30 in: Fungal Genomics, Applied Mycology and Biotechnology Series. D. K. Arora and G. G. Khachatourians, eds. Elsevier Science, Burlington, MA.
15. Cesati, V. 1852. Klotzsch, herbarium vivum mycologicum, sistens fungorum per totam Germanium cresentium collection perfectam. Cent. XVII. Flora 35:398.
16. Chen, F., Goodwin, P. H., Khan, A., and Hsiang, T. 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris*. Can. J. Microbiol. 48:427-436.
17. Couch, B. C., Fudal, I., Lebrun, M.-H., Tharreau, D., Valent, B., Kim, P., Notteghem, J.-L., and Kohn, L. M. 2005. Origins of host-specific populations of the rice blast pathogen, *Magnaporthe oryzae*, in crop domestication with subsequent expansion of pandemic clones on rice and weeds of rice. Genetics 170:613-630.
18. Couch, H. B. 1979. Is it anthracnose or is it wilt? The Greensmaster 15:3-6.
19. Crouch, J., Clarke, B. B., and Hillman, B. I. 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. 10:186-195.
20. Dale, J. L. 1963. Corn anthracnose. Plant Dis. Rep. 47:245-249.
21. Darlu, P., and Lecointre, G. 2002. When does the incongruence length difference test fail? Mol. Biol. Evol. 19:432-437.
22. Demoeden, P. H. 2000. Stresses and maladies of creeping bentgrass. Pages 44-48 in: Creeping Bentgrass Management: Summer Stresses, Weeds and Selected Maladies. Ann Arbor Press, Chelsea, MI.
23. Dolphin, K., Belshaw, R., Orme, C. D. L., and Quicke, D. L. J. 2000. Noise and incongruence: Interpreting results of the incongruence length difference test. Mol. Phylogenet. Evol. 17:401-406.
24. Dowton, M., and Austin, A. D. 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy—the behavior of the incongruence length difference test in mixed-model analysis. Syst. Biol. 51:19-31.
25. Du, M., Schardl, C. L., Nuckles, E. M., and Vaillancourt, L. 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Colletotrichum* species complexes. Mycologia 97:641-658.
26. Fang, G. C., Hanau, R. M., and Vaillancourt, L. J. 2002. The *SOD2* gene, encoding a manganese-type superoxide dismutase, is up-regulated during conidiogenesis in the plant-pathogenic fungus *Colletotrichum graminicola*. Fungal Genet. Biol. 36:155-165.
27. Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution 39:783-791.
28. Freeman, S. 2000. Genetic diversity and host specificity of *Colletotrichum* species on various fruits. Pages 131-144 in: *Colletotrichum*: Host Specificity, Pathology, and Host–Pathogen Interaction. D. Prusky, S. Freeman, and M. B. Dickman, eds. The American Phytopathological Society, St. Paul, MN.
29. Hey, J., and Wakeley, J. 1997. A coalescent estimator of the population recombination rate. Genetics 145:833-846.
30. Horvath, B. J., and Vargas, J. M. 2004. Genetic variation among *Colletotrichum graminicola* isolates from four hosts using isozyme analysis. Plant Dis. 88:402-406.
31. Hsiang, T., and Goodwin, P. H. 2001. Ribosomal DNA sequence comparisons of *Colletotrichum graminicola* from turfgrasses and other hosts. Eur. J. Plant Pathol. 107:593-599.
32. Hudson, R. R., and Kaplan, N. L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111:147-164.
33. Huelsenbeck, J. P., and Rannala, B. 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitutional models. Syst. Biol. 53:904-913.
34. Huelsenbeck, J. P., and Ronquist, F. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572-1574.
35. Huson, D. H. 1998. SplitsTree: Analyzing and visualizing evolutionary data. Bioinformatics 14:68-73.
36. Jamil, F. F., and Nicholson, R. L. 1987. Susceptibility of corn to isolates of *Colletotrichum graminicola* pathogenic to other grasses. Plant Dis. 71:809-810.
37. Johnston, P. R., and Jones, D. 1997. Relationships among *Colletotrichum* isolates from fruit-rots assessed using rDNA sequences. Mycologia 89:420-430.
38. Kemp, G. H. J., Pretorius, Z. A., and Smith, J. 1991. Anthracnose of wheat in South Africa. Phytophylactia 23:177-179.
39. Larget, B., and Simon, D. L. 1999. Markov chain Monte Carlo algorithms for the Bayesian analysis of phylogenetic trees. Mol. Biol. Evol. 16:750-759.
40. LeBeau, F. J. 1950. Pathogenicity studies with *Colletotrichum* from different hosts on sorghum and sugarcane. Phytopathology 40:430-438.
41. LeBeau, F. J., Stokes, I. E., and Coleman, O. H. 1951. Anthracnose and red rot of sorghum. Vol. 1035. U.S. Dep. Agric., Washington, D.C.
42. Liyanage, H. D., McMillan, R. T., and Kistler, H. C. 1992. Two genetically distinct populations of *Colletotrichum gloeosporioides* from citrus. Phytopathology 82:1371-1376.
43. Messiaen, C. M., Lafon, R., and Malot, O. 1959. Necroses de racines, pourritures de tiges et verse parasitaire du Mais. Ann. Epiphyt. Ser. C. 10:441-474.
44. Minussi, E., and Kimati, H. 1979. Taxonomy of *Colletotrichum graminicola* (Ces.) Wils. (sensu Arx, 1957). Rev. Centro Cienas Rurais 9:171-187.
45. Moriwaki, J., Tsukiboshi, T., and Sato, T. 2002. Grouping of *Colletotrichum* species in Japan based on rDNA sequences. J. Gen. Plant Pathol. 68:307-320.
46. O'Donnell, K., and Cigelnik, E. 1997. Two divergent intragenomic rDNA ITS2 types within a monophyletic lineage of the fungus *Fusarium* are nonorthologous. Mol. Phylogenet. Evol. 7:103-116.
47. Posada, D., and Crandall, K. A. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817-818.
48. Posada, D., and Crandall, K. A. 2001. Intraspecific gene genealogies: Trees grafting into networks. Trends Ecol. Evol. 16:37-45.
49. Posada, D., and Crandall, K. A. 2002. The effect of recombination on the accuracy of phylogeny estimation. J. Mol. Evol. 54:396-402.
50. Posada, D., Crandall, K. A., and Holmes, E. C. 2002. Recombination in evolutionary genomics. Annu. Rev. Genet. 36:75-97.
51. Price, E. W., and Carbone, I. 2005. SNAP: Workbench management tool for evolutionary population genetic analysis. Bioinformatics 21:402-404.
52. Randhir, R. J., and Hanau, R. M. 1997. Size and complexity of the nuclear genome of *Colletotrichum graminicola*. Appl. Environ. Microbiol. 63:4001-4004.
53. Rozas, J., Sanchez-DelBarrio, J. C., Messeguel, X., and Rozas, R. 2003. DnaSP, DNA sequence polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.
54. Selby, A. D., and Manns, T. F. 1909. Studies in diseases of cereals and grasses. Ohio Agric. Exp. Stn. Bull. 203:207.
55. Sherriff, C., Whelan, M. J., Arnold, G. M., and Bailey, J. A. 1995. rDNA sequence analysis confirms the distinction between *Colletotrichum graminicola* and *C. sublineolum*. Mycol. Res. 99:475-478.
56. Sreenivasaprasad, S., Mills, P. R., Meehan, B. M., and Brown, A. E. 1996. Phylogeny and systematics of 18 *Colletotrichum* species based on ribosomal DNA spacer sequences. Genome 39:499-512.
57. Sullivan, J. 1996. Combining data with different distributions of among-site rate variation. Syst. Biol. 45:375-380.
58. Sutton, B. C. 1965. Studies on the taxonomy of *Colletotrichum* Cda with especial reference to *C. graminicola* (Ces.) Wilson. University of London, London.
59. Sutton, B. C. 1966. Development of fruitifications in *Colletotrichum graminicola* (Ces.) Wils. and related species. Can. J. Bot. 44:887-897.
60. Sutton, B. C. 1968. The appressoria of *Colletotrichum graminicola* and *C. falcatum*. Can. J. Bot. 46:873-876.
61. Sutton, B. C. 1980. The coelomycetes: Fungi imperfecti with pycnidia, acervuli, and stromata. Commonwealth Mycological Institute, Kew, UK.
62. Sutton, B. C. 1992. The genus *Glomerella* and its anamorph *Colletotrichum*. Pages 1-26 in: *Colletotrichum*: Biology, Pathology and Control. J. A. Bailey and M. J. Jeger, eds. CAB International, Wallingford, UK.
63. Swofford, D. L. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer Associates, Sunderland, MA.
64. Taylor, J. W., Jacobson, D. J., Kroken, S., Kasuga, T., Geiser, D. M., Hibbett, D. S., and Fisher, M. C. 2000. Phylogenetic species recognition and species concepts in fungi. Fungal Genet. Biol. 31:21-32.

Appendix 2: Publication version of Chapter 1

65. Thompson, J. D., Higgins, D. G., and Gibson, T. J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. Nucleic Acids Res. 22:4673-4680.
66. Vaillancourt, L., Du, M., Wang, J., Rollins, J., and Hanau, R. 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia 92:430-435.
67. Vaillancourt, L. J., and Hanau, R. M. 1991. A method for genetic analysis of *Glomerella graminicola* (*Colletotrichum graminicola*) from maize. Phytopathology 81:530-534.
68. Vaillancourt, L. J., and Hanau, R. M. 1992. Genetic and morphological comparisons of *Glomerella* (*Colletotrichum*) isolates from maize and sorghum. Exp. Mycol. 16:219-229.
69. Valero, H. M., Resende, M. A., Weikert-Oliveira, R. C. B., and Casela, C. R. 2005. Virulence and molecular diversity in *Colletotrichum graminicola* from Brazil. Mycopathologia 159:449-459.
70. White, T. J., Bruns, T., Lee, S., and Taylor, J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. Pages 315-322 in: PCR Protocols: A Guide to Methods and Applications. M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, eds. Academic Press, San Diego.
71. Wilson, G. W. 1914. The identity of the anthracnose of grasses in the United States. Phytopathology 4:106-112.
72. Zeiders, K. E. 1987. Leaf spot of Indiangrass caused by *Colletotrichum caudatum*. Plant Dis. 71:348-350.
73. Zwillenberg, H. H. L. 1959. *Colletotrichum graminicola* (Ces.) Wilson auf mais und verschiedenen anderen Pflanzen. Phytopathol. Z. 34:417-425.

Appendix 3: Publication version of Chapter 2

Available online at www.sciencedirect.com
ScienceDirect
Fungal Genetics and Biology 45 (2008) 190–206
FUNGAL GENETICS AND BIOLOGY
www.elsevier.com/locate/yfgbi

The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen

Jo Anne Crouch, Bernadette M. Glasheen, Michael A. Giunta, Bruce B. Clarke, Bradley I. Hillman *

Department of Plant Biology and Pathology, Rutgers University, 59 Dudley Road, New Brunswick, NJ 08901-8520, USA

Received 8 April 2007; accepted 3 August 2007
Available online 29 August 2007

Abstract

Mobile transposable elements are among the primary drivers of the evolution of eukaryotic genomes. For fungi, repeat-induced point mutation (RIP) silencing minimizes deleterious effects of transposons by mutating multicopy DNA during meiosis. In this study we identify five transposon species from the mitosporic fungus *Colletotrichum cereale* and report the signature pattern of RIP acting in a lineage-specific manner on 21 of 35 unique transposon copies, providing the first evidence for sexual recombination for this species. Sequence analysis of genomic populations of the retrotransposon *Ccret2* showed repeated rounds of RIP mutation acting on different copies of the element. In the RIPped *Ccret2* population, there were multiple inferences of incongruence primarily attributed to RIP-induced homoplasy. This study supports the view that the sequence variability of transposon populations in filamentous fungi reflects the activities of evolutionary processes that fall outside of typical phylogenetic or population genetic reconstructions.
© 2007 Elsevier Inc. All rights reserved.

Keywords: *Colletotrichum*; Transposon; Retrotransposon; Repeat-induced point mutation; Recombination; Sex; Evolution; Homoplasy

1. Introduction

Mobile genetic elements such as transposons (TEs) are abundant in eukaryotes, and with the exception of *Plasmodium falciparum* (Gardner et al., 2002), the causative agent of human malaria, TEs populate the DNA of all well-studied organisms. TEs may occupy a substantial proportion of the host genome: 60% of the maize genome is transposon-derived (Messing and Dooner, 2006), as is 38% of the mouse genome (IMGSC, 2001) and 45% of the human genome (IHGSC, 2001). In contrast, the genomes of many eukaryotes are composed of relatively few transposons: for example, only 4.3% of the chicken genome is transposon-derived (Wicker et al., 2005). A relatively small contribution of TEs to the genomes of fungi is typical, with only 3.1% of the *Saccharomyces cereviseae* genome comprised of TEs (Goffeau et al., 1996), while 8.2–14% of the genome of the rice blast fungus, *Magnaporthe oryzae* may be derived from TEs (Dean et al., 2005; Thon et al., 2006).

Because TEs are able to move about the host genome and insert into a host's DNA through either cut-and-paste (DNA, or Class II transposons) or copy-and-paste mechanisms via RNA intermediates (retro, or Class I transposons), these elements can exert a significant influence on the fitness and evolutionary potential of their hosts through events such as insertional mutagenesis, disrupted or enhanced gene expression or gross chromosomal rearrangements (Hua-Van et al., 2005). Given the numerous ways that transposition can impact the genome, a variety of methods have evolved to safeguard the host against the effects of potentially deleterious insertions or unsupportable transposition rates. In several organisms, highly specific targeting mechanisms have been shown to limit

* Corresponding author. Fax: +1 732 932 9377.
E-mail address: hillman@aesop.rutgers.edu (B.I. Hillman).

1087-1845/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.fgb.2007.08.004

Appendix 3: Publication version of Chapter 2

TE integration to non-essential genomic regions, thereby protecting host integrity. Most TEs appear to have integration "hotspots" that are dictated by nucleotide sequence, patterns of hydrogen bonds, DNA-bending proteins and/or DNA conformation (Chalmers et al., 1998; Bender and Kleckner, 1992; Ketting et al., 1997; Liu et al., 2005). Well known examples of targeted integration in fungi include the Ty retroelements of *S. cerevisiae*, which insert preferentially upstream of pol III transcribed genes and in silent chromatin regions (Zou et al., 1996; Devine and Boeke, 1996; Chalker and Sandmeyer, 1992) and the retrotransposon Tf1 in *Schizosaccharomyces pombe*, which exhibits a clear preference for integration in tandem and divergent intergenic pol II promoter regions (Singleton and Levin, 2002).

Filamentous fungi actively regulate repetitive sequences through silencing mechanisms such as quelling (RNA silencing) (Cogoni et al., 1996), meiotic silencing (Shiu et al., 2001), and repeat-induced point mutation (RIP) (Cambareri et al., 1989). The RIP mutation process is remarkably efficient in disabling transposable elements through the detection and subsequent mutation of duplicated sequences longer than ~400 bp (Watters et al., 1999). Just prior to karyogamy, GC-to-AT transitions are induced in duplicate sequences sharing >80% similarity, with as many as 30% of GCs converted to ATs (Cambareri et al., 1989) and repetitive DNA remaining susceptible to "RIPping" through six generations (Cambareri et al., 1991). Since its initial discovery in *Neurospora crassa* (Selker et al., 1987), the RIP-mutation process has been identified experimentally or through sequence analysis in the ascomycetes *Aspergillus fumigatus* (Neuveglise et al., 1996), *Aspergillus nidulans* (Nielson et al., 2001; Clutterbuck, 2004), *Aspergillus oryzae* (Montiel et al., 2006), *Fusarium oxysporum* (Hua-Van et al., 2001), *Leptosphaeria maculans* (Attard et al., 2005), *M. oryzae* (Nakayashiki et al., 1999; Ikeda et al., 2002), *N. tetrasperma* (Bhat et al., 2004); *Ophiostoma* sp. (Bouvet et al., 2007) and *Podospora anserina* (Graia et al., 2001) and in the basidiomycete *Microbotryum violacum* (Hood et al., 2005), although RIP-mutation activity in these species has always been found in a much less aggressive form than that observed in *N. crassa* (Galagan and Selker, 2004).

We have been developing several molecular tools, including transposon-based marker systems, to increase our understanding of the recent emergence of the mitosporic ascomycete fungus *Colletotrichum cereale* as a pathogen of turfgrasses and its benign existence in natural grassland and agroecosystems. Beginning in the mid 1990s, *C. cereale* emerged as one of the most destructive pathogens of cool-season turfgrasses (Smiley et al., 2005). Outside of golf course greens, *C. cereale* is a common inhabitant of a wide range of C3 cereals and grasses of the grass subfamily Pooideae, where it survives without inducing noticeable levels of disease (Crouch et al., 2006, J.A. Crouch and B.I. Hillman, unpublished data). Little is known about *C. cereale* populations and, until recently, the fungus was thought to be conspecific with *C. graminicola*, a pathogen of corn (Crouch et al., 2006). Two major *C. cereale* lineages (clades A and B) have been described using sequences of the internal transcribed spacer (ITS) region of the ribosomal DNA (Crouch et al., 2005) and multilocus phylogenetic analyses (Crouch et al., 2006), but the evolutionary processes that shaped these lineages remain largely unexplored.

During the course of surveying the *C. cereale* genome for TEs suitable for use as molecular markers, we observed that many of this organism's transposon sequences were distinguished by a pronounced A + T nucleotide bias; subsequent bioinformatics analysis demonstrated this bias reflected the characteristic patterns of RIP-like C → T and G → A transitions. In this study, five different species of transposons were identified from the two major lineages of *C. cereale* in RIPped and "normal", non-mutated forms: two DNA transposons, two species of long-terminal repeat (LTR) retrotransposons and one non-LTR retrotransposon. In this paper, we describe these *C. cereale* transposable elements and document how the process of RIP mutation has considerably altered 21 of 35 unique transposon copies surveyed in a lineage-specific manner. We then employ the *Ccret2* retrotransposon *pol* gene sequence to explore the impact of RIP-mutated transposons when these elements are used to generate evolutionary inferences for phylogenetic and population genetic analyses.

2. Materials and methods

2.1. Construction of genomic DNA libraries and the identification of repetitive transposon DNA

Genomic DNA was isolated from fungal mycelia using a standard phenol:chloroform extraction protocol (Sambrook and Russell, 2001). *C. cereale* genomic DNA libraries were constructed from the EcoRI-digested DNA of isolates PA-50231 (clade A) and PA-50005 (clade B) in the plasmid vector pGEM3zf+ (Promega, Madison, WI); the culture and origin of these isolates was described previously (Crouch et al., 2006). To screen for repetitive sequences, insert-bearing colonies were transferred to Colony/Plaque Screen Transfer Membranes (NEN, Boston, MA) following the manufacturer's alkaline lysis protocol. The Random Primers DNA Labeling System (Invitrogen Corp., Carlsbad, CA) was used to label *C. cereale* genomic DNA with [α^{32}P]dCTP (MP Biomedicals, Irvine, CA). After a 30-min prehybridization in hybridization solution [7% SDS, 1 mM EDTA and 0.5 M Na_2HPO_4] at 65 °C, the membranes were hybridized overnight with the denatured, labeled total genomic DNA probe in fresh hybridization solution at 65 °C. Two high stringency washes were performed at 65 °C [5% SDS, 1 mM EDTA and 40 mM Na_2HPO_4] for 20 min per wash; followed by two additional washes [1% SDS, 1 mM EDTA and 40 mM Na_2HPO_4] for 20 min each. The hybridized membranes were exposed to autoradiography film (Lab Scientific Inc., Livingston, NJ)

Appendix 3: Publication version of Chapter 2

in the presence of an intensifying screen and incubated at −70 °C for 24–48 h before development.

Plasmid DNA from clones containing putative repetitive DNA inserts, as predicted by the occurrence of significant levels of hybridization, was purified using the QIAprep Spin Miniprep Kit (Qiagen Inc., Valencia, CA).

To explore the potential distribution of RIPped TEs in *C. cereale* clade A, an ordered cosmid library of 6144 clones was constructed in the pWEB vector (EPICENTRE Biotechnologies, Madison, WI) from the genomic DNA of the *C. cereale* species epitype strain NJ-6340 (clade A) (Crouch et al., 2006). Colonies were transferred to Zeta-probe membranes (Bio-Rad Laboratories, Hercules, CA) using the alkaline lysis protocol and probed with [α^{32}P]dCTP-labelled amplicon of the RIP-mutated *C. cereale* retrotransposon *Ccret1*^{12IIP6} as described above. High quality plasmid purifications were prepared from the strongly hybridizing cosmid clone 9F8 using the Nucleobond AX Plasmid Maxi kit (BD Biosciences, Easton, PA).

2.2. Sequence analysis

Insert DNA was sequenced in both directions from the T7 and SP6 primer sites using the ABI Prism BigDye sequencing chemistry (Applied Biosystems Inc., Foster City, CA) on ABI 3100 capillary sequencer. Sequence data was used to perform BLAST searches (Altschul et al., 1990) against the National Center for Biotechnology Information (NCBI) GenBank database to identify sequences with similarity to transposable elements.

Multiple sequence alignments of the putative *C. cereale* transposons and similar transposons identified from GenBank for phylogenetic analyses were generated using Clustal W (Thompson et al., 1994) as launched in MegAlign (DNASTAR Inc., Madison, WI), manually adjusted to exclude gaps and ambiguously aligned regions from the dataset. Amino acid alignments were evaluated using maximum likelihood (ML) analyses run in PHYLIP v3.66 (Felsenstein, 2006) modeled under the best fit evolutionary parameters obtained using ProtTest v1.3 (Abascal and Posada, 2005). Posterior probabilities supporting the ML tree topologies were generated using Mr. Bayes v3.1 (Huelsenbeck and Ronquist, 2003) by performing two simultaneous runs of one cold and three heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC) for 10,000,000 generations and sampling trees every 500 generations. Fixed-rate evolutionary models for the amino acid data were estimated by MCMCMC model jumping between nine models of protein evolution; each model contributed to the posterior distribution according to the proportion of its posterior probability. Run diagnostics were performed every 1000th generation, with the first 25% of the output discarded as burn-in. To confirm that the two separate runs converged to a stationary distribution, the average standard deviation of split frequencies, the potential scale reduction factor convergence diagnostic, and the plot of generation versus log likelihood values were each examined for values and distributions characteristic of samples drawn from a posterior probability distribution. Trees sampled from the posterior distribution were imported into PAUP* v.4.0b10 (Swofford, 2000) and used to construct 50% majority-rule consensus trees from which support values were derived.

2.3. Estimates of repeat-induced point mutation

To evaluate patterns of repeat-induced point mutation in the transposons identified in this study, 110 DNA sequences (GenBank Accession Nos. DQ663091–DQ663135, DQ666147–DQ666164 and DQ667983–DQ668029) from the *C. cereale* genome were generated from several sources: random sequencing using the universal priming sites of clones from the PA-50005 and PA-50231 plasmid DNA libraries constructed for colony hybridizations (Section 2.1); sequencing from universal priming sites using a plasmid DNA library constructed from the isolate NJ-6340 in the vector pGEM-T Easy (Promega, Madison, WI); sequencing from universal priming sites using a cDNA library constructed in the vector pGEM-T Easy from *C. cereale* clade A isolate KS-20B-DGU; a 5-kb sequence from the mating type locus from the NJ-6340 cosmid library (Section 2.1); and direct sequencing from the PCR amplicons of isolate NJ-6340 genomic DNA from three single-copy genes (manganese superoxide dismutase, β-tubulin 2, and a class V chitin synthase; Table 1). After removing all ribosomal, mitochondrial and repetitive DNA sequences, a concatenated sequence of 70,594 bp was constructed from the 110 individual sequences to calculate RIP indices in a method modified from the protocol originally used to study RIP patterns in *N. crassa* (Margolin et al., 1998). The computer program Swaap (Pride, 2004) was used to identify observed and expected patterns of dinucleotides in a series of 200 bp sliding windows shifted in 100-bp intervals from the concatenated genomic sequence and the transposon sequences.

2.4. Estimation of incongruent evolutionary relationships caused by homoplasy or recombination

Since the retrotransposon *Ccret2* was found in multiple copies, with both RIPped and unmutated copies harbored in the *C. cereale* clade B genome, this element was used to determine whether RIPped sequence data could serve as reliable indicators of *C. cereale* evolutionary relationships. PCR products were amplified using the *Ccret2*^{A15} primer pair MV-Pol-2F/MV-Pol-3R (*Ccret2*$^{POL2/3}$, Table 1) from *C. cereale* isolates PA-50005, PA-50231 and *C. sublineolum* isolate S12001 genomic DNA. PCR products were gel purified using the Gene Clean III kit (Qbiogene, Irvine, CA), and cloned into the vector pGEM-T Easy (Promega, Madison, WI). Thirty-nine inserts from *Ccret2*$^{POL2/3}$ colonies were sequenced in both directions; the resultant nucleotide data were aligned as described above (Section 2.2) and analyzed using ML in PAUP*, with the closely related, non-RIPped retrotransposon *Cgret* from *C. gloeosporioides*

Appendix 3: Publication version of Chapter 2

Table 1
Primer sequences used in this study

Gene	Region	Library (lineage)	Lineages amplified	Primer name	Sequence (5′ → 3′)
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-POL-1R	5′ CGTACGGTCCATGCTCTG 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-POL-2F	5′ CAGAGCATGGACCGTACG 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-POL-3R	5′ CAGTACCTTGTGTATGTG 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A	MV-GAG-1R	5′ TCTGCATTCGTCGTAGAG 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A	MV-GAG-2F	5′ CTCTACGACGAATGCAGA 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A	MV-GAG-3F	5′ CGAGCAAAATCGAACGAA 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-RET-1F	5′ AAGGCTGCATTACACTACG 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-RET-2R	5′ CAGGCGTGGAGTTCTTT 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-GAG-5F	5′ AATCCTTAGTCTTTATGTTCT 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	B	MV-GAG-6R	5′ TTATTATTACGCTAGTTATTATTT 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-7F	5′ CACTACGAAGGCAAAGCACAC 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-8R	5′ TATCACATCCAAGCGTCCTATCT 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-11F	5′ GTGCTATAACTGTAAGAAGAT 3′
Ccret2 LTR retrotransposon	A15	PA-50231 (A)	A, B	MV-GAG-12R	5′ CAGTTGGCGTTTGTCGTT 3′
Ccret2 LTR retrotransposon	DBP16	PA-50005 (B)	B	MV-GAG-20F	5′ GCTTAGTAGTAAAGTTAAG 3′
Ccret2 LTR retrotransposon	DBP16	PA-50005 (B)	B	MV-GAG-21R	5′ CTGTAATGTTAAGTCTAG 3′
Ccret1 LTR retrotransposon	DBP6	PA-50005 (B)	B	PV-INT-40F	5′ AGGGCTGTGTCAATACTCA 3′
Ccret1 LTR retrotransposon	DBP6	PA-50005 (B)	B	PV-INT-41R	5′ GTCTTCCCTTCCACTGTTA 3′
Collect1 DNA transposon	I-29	PA-50005 (B)	B	*pogo*-20F	5′ GGTAGGTATGCCTTATAC 3′
Collect1 DNA transposon	I-29	PA-50005 (B)	B	*pogo*-21R	5′ CCTTCTAATACTTACTTAG 3′

(Zhu and Oudemans, 2000) used as the outgroup taxon. Starting trees were obtained from 10 random-addition replicates using a tree-bisection reconnection branch swapping algorithm. Posterior probabilities supporting the tree topology were generated using Bayesian inference as described above (Section 2.2). The model of evolution that best described the dataset was estimated using Modeltest v3.06 (Posada and Crandall, 1998) and specified for both the ML and Bayesian estimates.

Several independent analyses were performed on the *Ccret2*$^{POL2/3}$ multiple sequence alignment to detect incongruence due to recombination and discriminate its influence from the effects of homoplasy induced by the RIP mutation process. First, we graphically assessed the dataset for evidence of network-like relationships between the *Ccret2*$^{POL2/3}$ copies using the split decomposition method implemented in SplitsTree v4.2 (Huson, 1998; Huson and Bryant, 2006). Because a subset of the sequences was characterized by relatively high A + T nucleotide base composition, the split decomposition analysis was conducted using LogDet distance (Lockhart et al., 1994) to ensure groupings were the result of legitimate signal and not confounded by compositional biases. Next, to detect discordant relationships inferred along the length of the *Ccret2*$^{POL2/3}$ sequences, the program SlidingBayes v0.94 (Paraskevis et al., 2005), working together with MrBayes, was used to perform a Bayesian sliding window analysis of the *Ccret2*$^{POL2/3}$ alignment to identify any incongruent tree topologies. A sliding window of 50 nucleotides was run along the length of the alignment, with four MCMCMC chains run for 10^6 generations per window; a 75% consensus tree was then generated in PAUP* from the resultant 3800 trees for each of the 20 windows. The Shimodaira–Hasegawa likelihood ratio test (Shimodaira and Hasegawa, 1999), performed in PAUP* using 10,000 RELL bootstrap replicates, was employed to test whether discordant topologies in the 20 consensus trees were generated along the length of the nucleotide alignment. Third, recombination in the dataset was estimated through calculation of the Phi statistic (ϕ_w) (Bruen et al., 2006) as launched through SplitsTree. Finally, possible recombination breakpoints in the multiple sequence alignment were assessed using GARD (Kosakovsky Pond et al., 2006).

2.5. Nucleotide sequence accession numbers

All new sequences generated by this study have been deposited in the GenBank database (Accession Nos.: DQ663091–DQ663113, DQ663495–DQ663534, DQ666106–DQ666136, DQ666147–DQ666164, DQ667983–DQ668029 and EF067890–EF067895).

3. Results

3.1. Identification and nomenclature of TEs from the *C. cereale* and *C. sublineolum* genomes

A total of 35 unique transposon copies were identified using a combination of methods: (1) five plasmid clones were identified as containing repetitive elements due to their strong hybridization signal when probed with *C. cereale* total genomic DNA; (2) 13 TEs were found during the course of random sequencing of the PA-50005 genomic DNA library; (3) six TEs were sequenced from subclones of the cosmid 9F8 identified through colony hybridization; and (4) nine *C. cereale* and two *C. sublineolum* retrotransposon sequences were obtained through PCR amplification and subsequent cloning. Significant levels of sequence similarity (Sections 3.2 and 3.3) strongly supported the identification of these elements as transposons. To simplify

Appendix 3: Publication version of Chapter 2

discussion in this narrative, the following names are used to describe the five new *C. cereale* TE species and one new *C. sublineolum* TE species identified in this study: (1) *Collect1* (*Colletotrichum cereale transposon 1*) is assigned to the *pogo* superfamily DNA transposon species and (2) *Collect2* (*Colletotrichum cereale transposon 2*) is used to describe the *Tc1/mariner* family TE. Retrotransposon taxa are named in accordance with the International Committee on Taxonomy of Viruses (ICTV) Code (ICTV, 2004): (1) *Colletotrichum cereale Ccret1 virus* (*Ccret1* (*Colletotrichum cereale retrotransposon 1*)) to describe the *C. cereale Pseudoviridae* family long-terminal repeat (LTR) retrotransposon; (2) *Colletotrichum cereale Ccret2 virus* (*Ccret2* (*Colletotrichum cereale retrotransposon 2*)) to describe the *C. cereale Metaviridae* family LTR retrotransposon; (3) *Colletotrichum cereale Ccret3 virus* (*Ccret3* (*Colletotrichum cereale retrotransposon 3*)) to describe the *C. cereale* non-LTR retrotransposon; and (4) *Colletotrichum sublineolum Cosret1 virus* (*Cosret1* (*Colletotrichum sublineolum retrotransposon 1*)) to describe the *C. sublineolum Metaviridae* family LTR retrotransposon. Individual copies of the TE species will be designated using superscript notation (e.g. *Ccret2*^{A15}).

3.2. Degenerate transposons are identified from the C. cereale clade B genome

Using a combination of BLASTX, PSI-BLAST and Conserved Domain Database (CDD) searches (Marchler-Bauer et al., 2005; Marchler-Bauer and Bryant, 2004), 24 sequences from the PA-50005 clade B genome were identified as members of three distinct transposon families (Table 2): the *pogo* superfamily DNA transposon *Collect1* (Fig. 1), the *Pseudoviridae* family retrotransposon *Ccret1* (Fig. 2), and the *Metaviridae* family retrotransposon *Ccret2* (Fig. 3). Of the 24 *C. cereale* TEs, 23 were predicted to be interrupted by at least one and as many as 52 stop codons, suggesting that these elements would not function autonomously. These 23 TEs were also notable in that they displayed a marked bias for A + T nucleotides, averaging 67.62% A + T in an organism with an estimated genome A + T content of 51.03% (Fig. 4a and Table 2). Despite the degeneracy of these TEs, several of the elements had recognizable conserved domains characteristic of transposons. The retrotransposon integrase core domain (gnl|CDD|25582), characteristic of *Pseudoviridae* retrotransposons, was present in both *Ccret1*^{DBP6} and *Ccret1*^{M31}, and the conserved DDE superfamily endonuclease domain required for DNA transposition (gi|CDD|26040) was identified from the sequence of *Collect1*^{M21}.

Although some of the *C. cereale* transposons of the same species were predicted to be present in overlapping regions as inferred by positional homology, reliable contigs could not be generated from the copies due to a low level of sequence similarity. Based upon their predicted position in the *gag* region, the nucleotide sequences of *Ccret2*^{4M13}, *Ccret2*^{1J24} and *Ccret2*^{1A1} would be expected to overlap with *Ccret2*^{DBP16} (Fig. 3), but the sequences were not similar enough to suggest derivation from a single genomic locus (>80% similarity). Similarly, despite positional homology, *Ccret1*^{M6}, *Ccret1*^{M56} and *Ccret1*^{7I1I4} did not possess enough sequence similarity to be drawn from a single TE copy. *Ccret1*^{M31} was 76% identical to *Ccret1*^{DBP6} over a 331-bp overlap, but only 49% of the differences between the two sequences were caused by transitions (Fig. 2). Similarly, *Collect1*^{M40} was predicted to share positional homology with *Collect1*^{I29} based on its alignment with *M. oryzae* transposases (Fig. 1), but the low level of sequence similarity (52%) showed that the two elements were not derived from the same TE copy. Thus, we predicted that although these elements are part of the same transposon species, these TEs clearly do not represent a single contiguous transposon sequence as evidenced by both differences in nucleotide base composition and the relatively low levels of sequence similarity in regions expected to overlap.

3.3. Intact transposon sequences are recovered from the C. cereale clade A genome

In contrast with the degenerate sampling of TEs recovered from the genome of the clade B isolate PA-50005, deduced translation products of the nine transposons from the *C. cereale* clade A genome (isolates PA-50231 and NJ-6340) were intact, with none interrupted by stop codons, and nucleotide compositions close to those observed in the *C. cereale* genome (Fig. 4a). Four transposon families were represented in our survey of the clade A genome: *Collect2*, a DNA transposon of the *Tc1/mariner* family, the LTR-retrotransposons *Ccret1* and *Ccret2*, and the non-LTR retrotransposon *Ccret3* (Table 2).

Because all of the transposon sequences from the *C. cereale* clade A genome were quite different from those found in the context of the clade B genome, we sought to determine whether there were clade B-like degenerate elements in the clade A genome. PCR primers designed to amplify the region between clade B's *Ccret2*^{DBP16} and the clade A sequence *Ccret2*^{A15} (Table 1 and Fig. 3) led to amplification of a 4478-bp product from the clade B isolate PA-50005 (*Ccret2*$^{Pol3/Gag11F}$), but no product in any clade A isolate, even at lower stringency annealing temperatures (data not shown), was found. Sequencing the cloned PCR amplicon revealed that *Ccret2*$^{Pol3/Gag11F}$ was characterized by numerous stop codons and deletions, similar to other TEs from the clade B genome.

In a second attempt to identify a clade B-like TE from a *C. cereale* clade A strain, the clade B *Ccret1*^{DBP6} sequence was used as a probe to screen an ordered cosmid library of clade A strain NJ-6340 (~3× coverage). Of 6,144 clones, only a single cosmid (9F8) hybridized strongly with the probe. Because clean sequence reads could not be obtained by direct sequencing the purified 9F8 cosmid DNA, the insert DNA was excised with *Not*I, then digested with

Appendix 3: Publication version of Chapter 2

Table 2
Dinucleotide pattern density of *C. cassiicola* sequence data

Transposon	RIP?	Isolate (clade)	Length (bp)	A+T%	RIP indices			Dinucleotide base frequencies (observed/expected)														Accession
					TA/AT	CA+TG/AC+GT	GC/CG	AA	TT	AC	TG	AG	TC	AT	TA	CA	GT	CC	GG	CG	GC	CT	GA	
Gulliver/ M40	—	PA-90005 (B)	669	42.91	0.58	0.59	0.46	0.71	0.55	1.26	0.57	0.62	1.76	1.01	0.59	0.83	1.12	0.73	0.75	1.50	0.70	1.03	1.38	DQ663507
Gulliver2	—	NJ-6340 (A)	1,465	46.65	0.87	1.17	1.03	1.06	1.36	0.95	0.99	0.98	0.94	0.91	0.79	0.85	0.62	0.95	0.88	1.06	1.09	1.12	1.16	EF067895
918-1465																								
Gulliver2	—	NJ-6340 (A)	1,256	49.88	0.83	1.69	3.02	0.74	0.68	0.62	1.11	1.20	0.89	1.38	1.14	1.16	0.72	1.39	1.14	0.41	1.24	1.13	0.86	EF067891
918-1256																								
Grem1	—	NJ-6340 (A)	1,787	45.80	0.65	1.16	0.92	1.03	1.23	0.84	1.02	1.02	1.05	1.07	0.70	0.93	0.84	1.09	0.86	1.06	0.97	0.82	1.17	EF067893
918-1787																								
Grem1	—	NJ-6340 (A)	662	43.35	0.76	1.13	0.65	1.05	1.05	0.89	0.94	1.00	1.04	1.09	0.83	0.83	0.68	0.94	0.68	1.17	0.76	1.01	1.21	EF067890
918-662																								
Grem2 M22	—	PA-90005 (B)	641	50.13	0.93	0.96	0.77	0.88	1.38	1.10	0.79	0.98	0.99	0.84	0.78	0.89	0.64	0.91	1.01	1.07	0.82	1.09	1.24	DQ663503
Grem2 A15	—	PA-90231 (A)	1,690	50.03	0.90	0.99	0.78	0.87	1.14	1.08	0.76	0.96	1.15	0.89	0.80	0.94	0.64	0.82	0.92	1.06	0.83	1.15	1.27	DQ663513
Grem2	—	PA-90231 (A)	461	51.38	0.89	0.96	0.76	0.86	1.13	1.07	0.78	0.99	1.22	0.88	0.78	0.94	0.72	0.84	0.96	1.10	0.84	1.15	1.26	DQ666317
PO12/3 MV2																								
Grem2 PO12/3 MV11	—	PA-90231 (A)	461	50.57	0.83	1.02	0.75	0.89	1.12	1.06	0.80	0.94	1.22	0.91	0.74	0.95	0.64	0.81	0.92	1.08	0.81	1.13	1.28	DQ666318
Grem2 PO12/3 MV25	—	PA-90005 (B)	461	50.13	0.93	0.96	0.77	0.88	1.38	1.10	0.79	0.98	0.99	0.84	0.78	0.89	0.64	0.91	1.01	1.07	0.82	1.09	1.24	DQ666310
Grem3	—	NJ-6340 (A)	2,137	46.57	0.80	1.16	1.08	1.05	1.11	0.88	0.89	1.10	1.23	0.89	0.71	0.98	0.73	1.16	0.73	0.95	1.03	1.18	1.09	EF067894
918-2137																								
Grem3	—	NJ-6340 (A)	1,558	48.26	0.57	1.18	1.02	0.85	1.11	1.01	1.16	1.02	1.12	0.98	0.56	1.17	0.96	0.91	0.82	0.94	0.96	0.92	1.20	EF067892
918-1558																								
Guest1/ PO12/3 MV23	—	S12001 (*C. sublineolum*)	461	51.57	0.91	0.98	0.74	0.93	0.46	1.08	0.81	0.97	1.15	0.89	0.81	0.93	0.70	0.88	0.95	1.08	0.79	1.14	1.31	DQ666306
Guest1/ PO12/3 MV29	—	S12001 (*C. sublineolum*)	461	52.95	0.83	1.46	0.99	0.90	1.25	1.13	1.27	0.97	1.04	0.91	0.57	0.99	0.41	0.81	0.91	0.86	0.85	1.40	1.43	DQ666313
Gulliver/ I29	RIP	PA-90005 (B)	587	70.98	2.11	0.27	1.47	0.94	0.92	1.27	0.23	1.66	0.58	0.65	1.37	0.37	1.00	1.28	0.90	0.46	0.68	1.81	1.07	DQ663509
Gulliver/ M21	RIP	PA-90005 (B)	630	66.10	2.11	0.36	17.38	0.65	0.57	1.24	0.47	1.71	0.43	0.80	1.69	0.39	1.13	1.49	1.02	0.05	0.95	1.84	0.78	DQ663502
Grem1/ DBR6	RIP	PA-50005 (B)	2,889	78.52	1.84	0.08	0.91	0.78	0.76	1.33	0.06	1.78	0.45	0.80	1.46	0.11	0.85	1.28	0.98	0.52	0.47	2.19	1.06	DQ663510
Grem1/ TH14	RIP	PA-90005 (B)	641	74.31	2.01	0.19	3.47	0.70	0.82	1.45	0.00	2.12	0.40	0.75	1.50	0.50	1.23	1.10	0.42	0.15	0.51	1.75	0.93	DQ663499
Grem1/ M6	RIP	PA-90005 (B)	624	68.19	2.21	0.32	4.75	0.76	0.83	0.97	0.25	1.92	0.76	0.69	1.53	0.38	0.98	1.00	1.22	0.25	1.21	1.83	0.65	DQ663501
Grem1/ M31	RIP	PA-90005 (B)	582	57.94	1.86	0.75	0.74	0.50	1.16	1.01	0.47	1.49	0.83	0.67	1.25	0.69	0.54	0.94	0.85	1.56	1.36	1.56	1.11	DQ663505
Grem1/ M56	RIP	PA-50005 (B)	418	73.94	1.80	0.23	1.96	0.89	0.79	0.83	0.36	1.76	1.11	0.79	1.42	0.13	1.33	1.08	1.21	0.41	0.79	1.68	0.41	DQ663508
Grem2 DBR16	RIP	PA-90005 (B)	2,380	67.01	2.19	0.55	1.82	0.78	1.17	1.39	0.45	1.52	0.36	0.59	1.29	0.68	0.67	0.80	0.69	0.71	1.30	1.69	1.06	DQ663511
Grem2 1A1	RIP	PA-50005 (B)	566	67.09	1.94	0.60	1.80	0.80	1.08	1.24	0.35	1.52	0.48	0.65	1.25	0.80	0.68	0.87	0.92	0.68	1.22	1.51	0.99	DQ663495
Grem2 4M13	RIP	PA-90005 (B)	535	66.67	1.92	0.56	2.55	0.80	1.19	1.22	0.26	1.62	0.40	0.67	1.29	0.81	0.68	0.92	0.98	0.55	1.39	1.49	0.89	DQ663498
Grem2 M3	RIP	PA-90005 (B)	689	69.62	1.90	0.39	1.34	0.80	1.03	1.25	0.13	1.58	0.48	0.72	1.36	0.64	0.73	1.04	1.46	0.54	0.94	1.63	0.97	DQ663500
Grem2 416	RIP	PA-90005 (B)	594	61.34	1.77	0.77	1.73	0.81	1.26	1.19	0.49	1.44	0.70	0.82	1.08	0.95	0.69	0.79	0.94	0.64	1.11	1.43	1.04	DQ663497

(continued on next page)

Appendix 3: Publication version of Chapter 2

Table 2 (continued)

Transposon	RIP?	Isolate (clade)	Length (bp)	A+T%	RIP indices			Dinucleotide base frequencies (observed/expected)														Accession
					TA/AT	CA+TG/AC+GT	GC/CG	AA	TT	AC	TG	AG	TC	AT	TA	CA	GT	CC	GG	CG	GC	CT	GA	
Ccret2 M35	RIP	PA-50005 (B)	192	70.41	2.04	0.25	1.73	0.75	0.73	1.21	0.45	1.71	0.92	0.78	1.59	0.18	1.24	0.73	1.71	0.39	0.67	1.77	0.31	DQ663506
Ccret2 1D24	RIP	PA-50005 (B)	619	71.34	2.26	0.20	8.73	0.80	0.86	0.75	0.24	2.23	0.90	0.66	1.50	0.12	1.10	1.09	0.93	0.16	1.44	1.87	0.50	DQ663496
Ccret2 Pol3/Gag11	RIP	PA-50005 (B)	4,478	65.97	1.93	0.58	2.89	0.87	0.99	1.15	0.54	1.53	0.61	0.70	1.35	0.65	0.91	1.06	1.13	0.38	1.10	1.67	0.85	DQ663512
Ccret2 POL2/3MV34	RIP	PA-50005 (B)	461	64.06	1.87	0.52	2.35	0.77	1.16	1.20	0.19	1.69	0.43	0.70	1.30	0.80	0.69	0.92	0.83	0.57	1.35	1.49	0.97	DQ666123
Ccret2 POL2/3 MV43	RIP	PA-50005 (B)	461	66.67	1.92	0.56	2.55	0.80	1.19	1.22	0.26	1.62	0.40	0.67	1.29	0.81	0.68	0.92	0.98	0.55	1.39	1.49	0.89	DQ666126
Ccret2 POL2/3 MV47	RIP	PA-50005 (B)	461	69.62	1.90	0.39	1.74	0.80	1.03	1.25	0.13	1.58	0.48	0.72	1.36	0.64	0.73	1.04	1.46	0.54	0.94	1.63	0.97	DQ666128
Ccret2 POL2/3 MV58	RIP	PA-50005 (B)	461	67.01	2.19	0.55	1.82	0.78	1.17	1.39	0.45	1.52	0.36	0.59	1.29	0.68	0.67	0.80	0.69	0.71	1.30	1.69	1.06	DQ666132
Ccret2 POL2/3 MV71	RIP	PA-50005 (B)	461	64.01	2.36	0.45	1.13	0.84	1.22	1.32	0.34	1.44	0.42	0.56	1.31	0.57	0.72	0.90	0.75	1.05	1.19	1.58	1.04	DQ666136
Ccret2 POL2/3 M57	RIP	PA-50005 (B)	461	61.34	1.77	0.77	1.73	0.81	1.26	1.19	0.49	1.44	0.70	0.61	1.08	0.95	0.69	0.79	0.94	0.64	1.11	1.43	1.04	DQ666131
Average, *C. cereale* genome			70,594	51.03	0.98	1.07	0.80	0.90	1.04	1.04	1.02	1.03	1.32	0.84	0.55	1.02	0.86	0.83	0.86	0.94	0.75	1.28	1.28	
Average, "normal" transposon sequences			14,170	45.32	0.79	1.10	0.88	0.91	1.09	1.00	0.91	0.99	1.12	0.95	0.74	0.96	0.73	0.95	0.89	1.02	0.90	1.09	1.21	
Average, RIP-mutated transposon sequences			19,210	67.62	2.00	0.44	1.93	0.78	1.05	1.19	0.31	1.66	0.58	0.68	1.36	0.56	0.85	0.99	1.00	0.55	1.06	1.67	0.88	

*Bam*HI and subcloned into plasmid pGEM3zf+. Six unique TE sequences drawn from three different species (*Ccret1*, *Ccret3* and *Collect2*) were identified on the cosmid insert. Two individual *Ccret1* sequences were also resident on the 9F8 cosmid, providing an explanation for the difficulties encountered during direct sequencing attempts. In contrast with the five *Ccret1* elements described from the clade B genome, both of the clade A-derived *Pseudoviridae* TEs possessed intact coding sequences and A + T compositions comparable to genomic levels (Fig. 4a).

Although none of the clade A TEs identified were degenerate, each of the nine sequences was unique (Figs. 2 and 3), suggesting that each represented a unique transposon copy. In particular, although *Ccret1*$^{9F8\text{-}1787}$ and *Ccret1*$^{9F8\text{-}662}$—sequences identified from a single cosmid clone—were predicted to overlap with each other and the other five *Ccret* copies through positional homology (Fig. 2), the sequences were too divergent to suggest a single TE copy. The clustering of two unique *Ccret1* elements with a copy of *Collect2* and *Ccret3* suggests that the 9F8 region of the *C. cereale* genome may serve as a target site for transposition.

3.4. Assessment of RIP-like patterns in the transposon sequences

The observation that 21 of the 35 TEs identified in this study—all from the *C. cereale* clade B genome—were characterized by high A + T and open-reading frame (ORF) interruption levels led us to hypothesize that the transposons in this lineage were being targeted and altered by the RIP mutation process (Cambareri et al., 1989). A comparative RIP index analysis (Margolin et al., 1998) supported this hypothesis. Overall, the genome was characterized by an A + T content of 51.03% with the ratio of all 16 possible dinucleotide combinations (observed/expected) found to span a range of 0.75–1.32 (Table 2, Figs. 4a and 5). In contrast, the dinucleotide sequence ratios in the high A + T clade B transposons were skewed in a manner indicative of RIP mutation. In all three of the *C. cereale* clade B transposon species, a clear pattern of increased TpA, CpT and ApG dinucleotides and decreased levels of TpG, CpA, CpG, GpA and TpC dinucleotides were detected (Table 2, Figs. 4a and 5) in sequences predicted to be RIPped based upon high A + T levels, but not in TEs with lower, genome-level A + T ratios.

The ratio of TpA to ApT dinucleotides, a signature of RIP in many fungal species (Margolin et al., 1998), in the concatenated *C. cereale* genomic DNA sequence data was 0.98, while the ratio for the predicted "normal", non-RIPped transposons was 0.79 (Table 2 and Fig. 4b). However, the TpA/ApT ratio for the predicted RIP-mutated clade B sequences was elevated relative to the frequencies observed in the *C. cereale* genome, with values ranging from 1.77 to 2.36, and 43% of these sequences characterized by ratios >2.0. These values are similar to those in RIPped elements in other fungal species (Fig. 4b), and sup-

Appendix 3: Publication version of Chapter 2

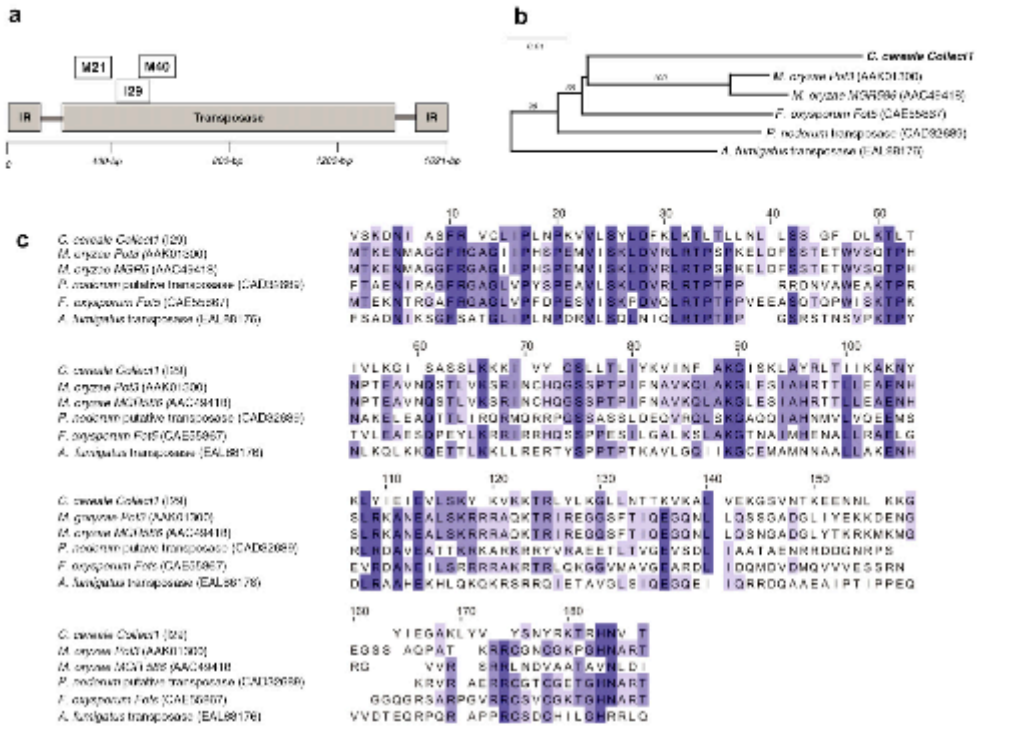


Fig. 1. The *pogo* family DNA transposon *Collect1*. (a) Diagram of the *Pot3* transposon from *Magnaporthe oryzae* showing the relative positions of the I29, M21 and M40 elements from *Colletotrichum cereale* isolate PA-50005. (b) Phylogenetic tree inferred from the amino acid alignment. Posterior probabilities are shown above the branches. (c) Amino acid alignment with putative homologues; dashes represent gaps in the alignment. Shading represents levels of conservation (dark shading, highly conserved position; no shading, poorly or non-conserved).

ported the identification of a *C. cereale* RIP mutation process.

To further test the hypothesis of RIP mutation, we employed the RIP index formula originally used in studies of *N. crassa* (Margolin et al., 1998) to compare the frequencies of the most common target of RIP-mutation in the *C. cereale* genome (CpA, CpG) to the frequencies of corresponding dinucleotide pairs of the same base composition (ApC, GpC). The ratio (CpA + CpG)/(ApC + GpC) for the *C. cereale* genome was 1.07, whereas the range for the "normal" *C. cereale* TEs averaged 1.10. In contrast, the (CpA + CpG)/(ApC + GpC) ratio for the predicted RIP-mutated transposons averaged 0.44, consistent with RIP-mutated TEs from several other ascomycetes (Table 2).

Filamentous fungi in which RIP-like mutations have been characterized show a clear preference for RIP-induced C → T transitions and complementary G → A transitions, with CpA, CpG and TpG the most common target sites in the ascomycetes *N. crassa*, *P. anserina*, *F. oxysporum*, *A. fumigatus*, *A. nidulans* and *L. maculans* (Attard et al., 2005; Cambareri et al., 1989; Clutterbuck, 2004; Graia et al., 2001; Hua-Van et al., 2001). Multiple sequence alignment of both RIPped and unRIPped homologues of *Ceret2*$^{POL2/3}$ copies (Section 3.5), revealed that cytosines in CpA and CpG and guanines in TpG are also likely targets of RIP in *C. cereale*. Of the transposon sequences predicted as RIP-mutated in this study, almost all showed decreased levels of these dinucleotides and elevated levels of the corresponding RIP-altered dinucleotide pairs (CpA → TpA; CpG → TpG → TpA; TpG → TpA) (Table 2 and Fig. 5). In comparison, the pattern of bases adjacent to RIP-mutated cytosines revealed no dominant pattern of site specificity, although the frequency of nucleotides found on the 3′-side of the mutated cytosine was A >>> T > C ~ G, while on the 5′-side the pattern was T > A > C > G. Together, these data provide considerable evidence that RIP mutation has acted upon transposons in *C. cereale*

Appendix 3: Publication version of Chapter 2

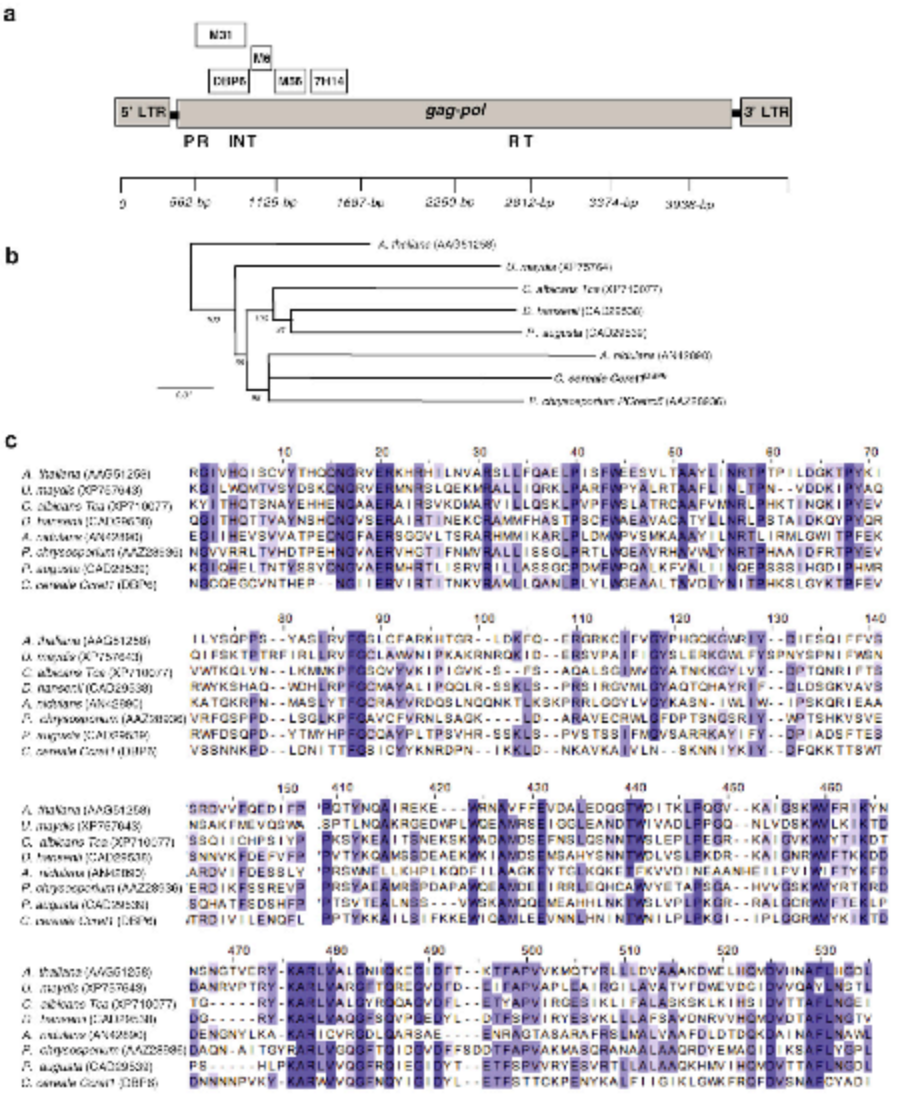


Fig. 2. The *Pseudoviridae* family DNA retrotransposon *Ceret1*. (a) Diagram of the *PCretro5* retrotransposon (AAZ28936) from *Phanerochaete chrysosporium* showing the relative positions of *Ceret1* elements M6, M31, M56, 7H14 and DBP6 from *Colletotrichum cereale* isolate PA-50005. (b) Phylogenetic tree inferred from the amino acid alignment. Posterior probabilities are shown above the branches. (c) Partial amino acid alignment with DBP6 and its putative homologues; dashes represent gaps in the alignment. Only the relatively conserved 234 residues used for the phylogenetic analysis are shown. Shading represents levels of conservation (dark shading, highly conserved position; no shading, poorly or non-conserved).

clade B; however, none of the clade A transposons identified in this study, including the 9F8 cosmid sequences that were originally detected by means of the RIPped *Ceret1*^{DBP6} probe sequence, showed any evidence of the RIP mutational process.

3.5. Genomic population analysis using Ceret2^{POL2I3} *sequence data*

Ceret2^{A15} is the only transposon in this study that is present in multiple copies in both of the major *C. cereale*

Appendix 3: Publication version of Chapter 2

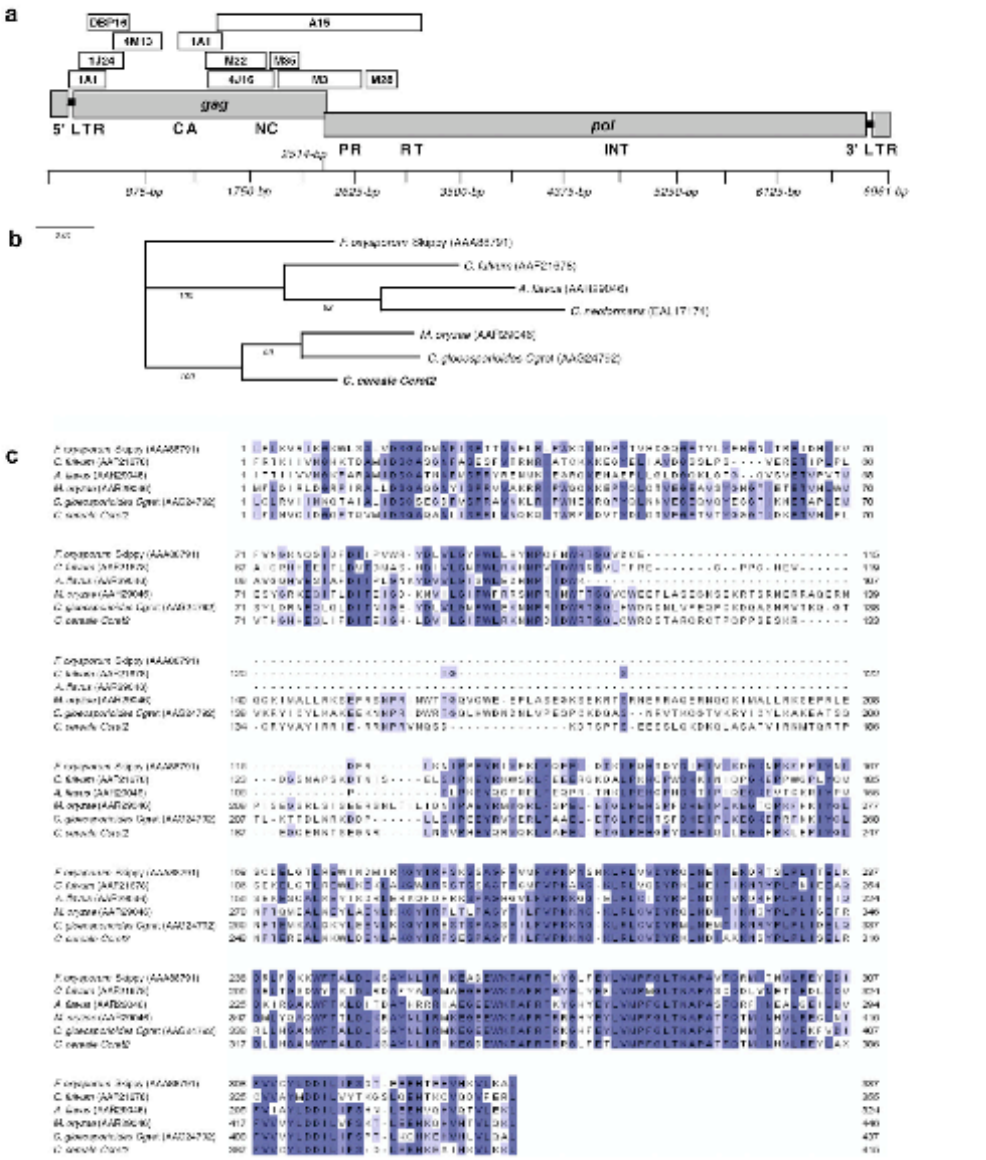


Fig. 3. The *Metaviridae* family retrotransposon *Ceret2*. (a) Diagram of the retrotransposon *C*gret (AAG24792) from *Colletotrichum gloeosporioides* showing the relative positions of *Ceret2* elements A15, DBP16, 4J16, M3, M22, 1A1, 1J24, 4M13, M28 and M35. (b) Phylogenetic tree inferred from the amino acid alignment. Posterior probabilities are shown above the branches. (c) Partial amino acid alignment with A15 and its putative homologues; dashes represent gaps in the alignment. Shading represents levels of conservation (dark shading, highly conserved position; no shading, poorly or non-conserved).

lineages (clades A and B) as determined by Southern blot analysis (data not shown). When the cloned, sequenced PCR product of $Ceret2^{A15}$ (“$Ceret2^{POL2/3}$”) was evaluated using the program Modeltest, the dataset was most accurately described by a general time reversible (GTR) model of evolution (GTR + G: A = 0.3594, C = 0.2328,

Appendix 3: Publication version of Chapter 2

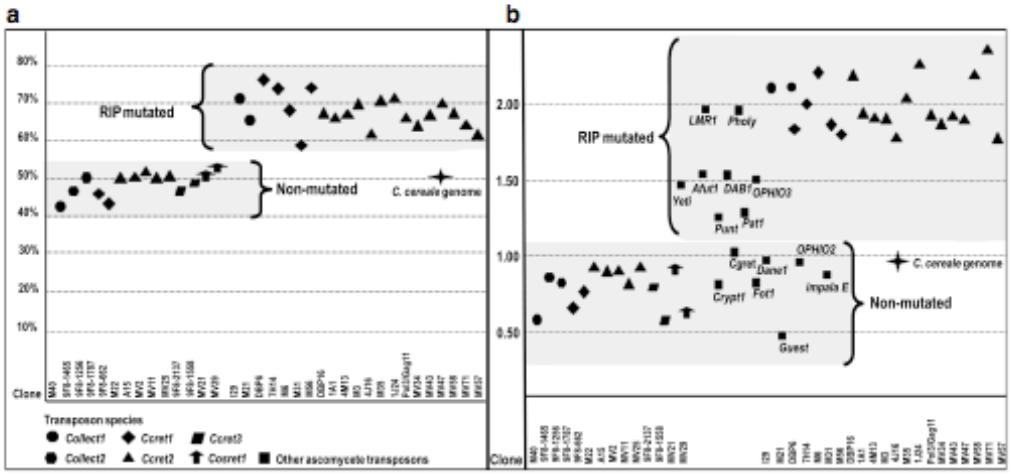


Fig. 4. (a) The *Colletotrichum cereale* transposon nucleotide base composition (A + T%). (b) Transposon TA/AT ratios for *C. cereale* and other ascomycete species. Accession numbers for additional RIPPED transposons: *Aspergillus nidulans Afut1*: L76086; *Fusarium oxysporum Fot1*_RIP: AF434909; *Leptosphaeria maculans Pholy*: AM084367; *L. maculans* LMR1: LM77515; *Neurospora crassa DAB1*: Y14976; *N. crassa Pant*: AF181821; *Ophiostoma ulmi OPHIO3*: DQ649005; *Podospora anserina Pat1*: AJ270953; *P. anserina Yeti*: AJ272171. Accession numbers for additional, non-mutated transposons: *A. nidulans Dane1*: AF295689; *C. gloeosporioides Cgret*: AF264028; *Cryphonectria parasitica Crypt1*: AF283502; *F. oxysporum impala E*: AF363407; *F. oxysporum Fot1*: X64799; *N. crassa Guest*: AY197334; *O. ulmi OPHIO2*: DQ649004.

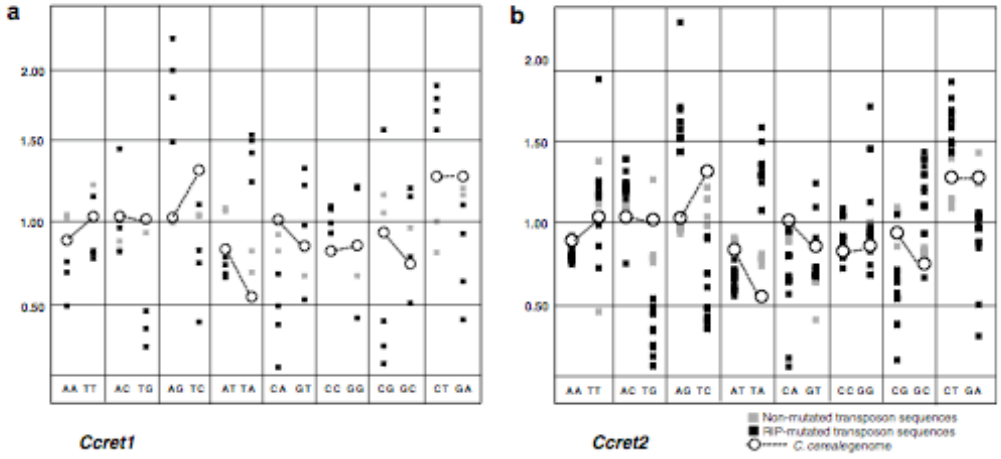


Fig. 5. Frequencies (expected/observed) of dinucleotide pairs in *Colletotrichum cereale* transposon sequences. (a) *Ccret1* and (b) *Ccret2*.

G = 0.1764, T = 0.2315; A → C 4.4901, A → G 1.8908, A → T 3.6501, C → G 3.7192; α = 1.1892), but partitioning of the dataset revealed that only the clade B lineage was accurately characterized by the time reversibility constraint; therefore, we used a non-GTR equivalent (HKY85 + G) to avoid overparameterization. Maximum likelihood phylogenetic analysis of the 461 bp *Ccret2*$^{POL2/3}$ alignment of 39 sequences resulted in a tree topology broadly congruent with *C. cereale* phylogenetic origin; both clade A and B were recovered (Fig. 6). In a notable exception, two sequences from the genome of *C. sublineolum* isolate S12001 (MV29 and MV21) formed a group with the *C. cereale* clade A *Ccret2*$^{POL2/3}$ elements, separate from the other nine *C. sublineolum* sequences (posterior probability = 100). Outside of this single exception, four main lineages were recovered: (1) the outgroup sequence, *Cgret*,

Appendix 3: Publication version of Chapter 2

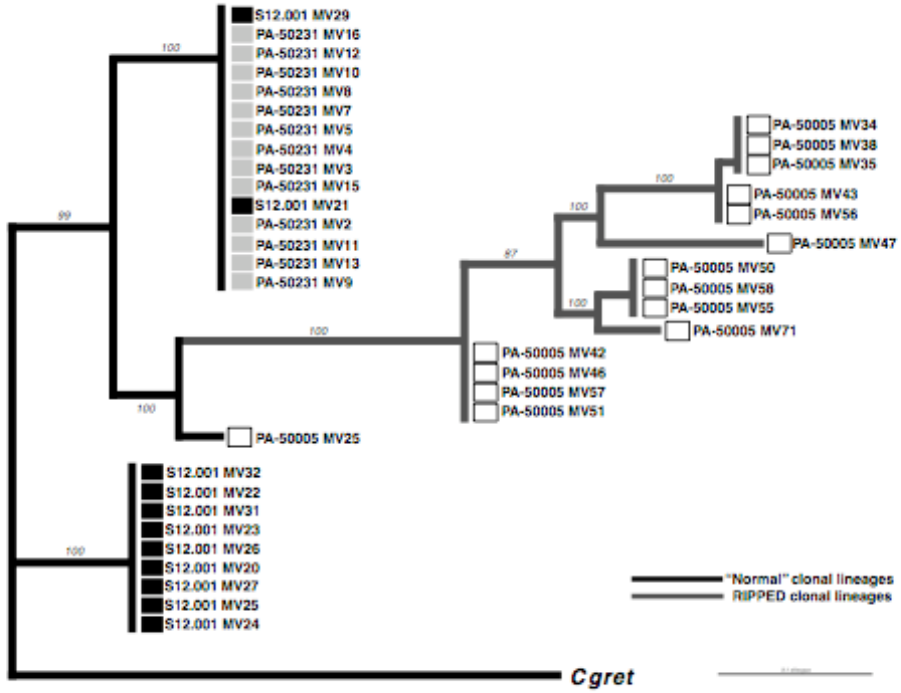


Fig. 6. *Ccret2*$^{POL2/3}$ phylogenetic analysis. The phylogenetic tree was constructed from cloned PCR amplicons (540-bp) from the *pol* region of *Ccret2*^{A15} in three *Colletotrichum* lineages: PA-50231 (14 clones, *Colletotrichum cereale* clade A), PA-50005 (15 clones, *C. cereale* clade B), and S12001 (11 clones, *C. sublineolum*); *Cgret* from *C. gloeosporioides* serves as the outgroup. The phylogenetic tree was generated from 75% consensus of 43,601 trees estimated using Bayesian phylogenetic inference.

from *C. gloeosporioides*; (2) the *C. sublineolum Cosret1* clade, which took the form of a polytomy; (3) the clade A-like *Ccret2*$^{POL2/3}$ lineage, consisting of all the PA-50231 *Ccret2*$^{POL2/3}$ copies and two sequences from *C. sublineolum* in a polytomy; and (4) the clade-B-derived *Ccret2*$^{POL2/3}$ lineage, which consisted of an extremely diverse, highly mutated assemblage of sequences, all from the genome of isolate PA-50005.

Evaluation of the *Ccret2*$^{POL2/3}$ sequences for A + T content, patterns of dinucleotides, TpA to ApT ratios and skewed frequency dinucleotide (CpA + CpG)/(ApC + GpC) ratios identified 14 of the sequences as RIP-mutated (Table 2), despite the fact that these sequences were generated using PCR primers designed from a non-RIPped sequence (*Ccret2*^{A15}). All 14 of the RIPped sequences were acquired from the genome of clade B isolate PA-50005; only a single copy from this isolate (MV25) was non-RIPped (Table 2) and placed as the basal lineage in the otherwise RIP-mutated clade B lineage. All transposon copies sequenced from the PA-50231 (clade A) and *C. sublineolum* S12001 genomes were "normal"; i.e., there was no evidence of RIP-like alterations to the sequences.

The *Ccret2*$^{POL2/3}$ dataset was evaluated for the presence of conflicting signal indicative of either homoplasy or recombination using four individual methods. Each of the analyses identified incongruence from the *C. cereale* clade B lineage, but the clade A and *C. sublineolum*-derived sequences were accurately depicted in a bifurcating tree topology consistent with the absence of both recombination and homoplasy. First, the split decomposition network visualized using LogDet distance was tree-like except among the RIPped clade B copies, where multiple incompatibilities were detected (Fig. 7); this analysis also confirmed that the clustering of taxa was the result of phylogenetic signal rather than nucleotide compositional bias. The second test, calculation of the Phi statistic (ϕ_w), confirmed the partitioned presence of recombination in the dataset and identified statistically significant levels of recombination in clade B both with and without the non-RIPped MV25 ($P = 0.02$ and 1.87E−4, respectively), but in all other clades, either individually or combined, and the dataset as a whole, the signature of recombination was not observed ($P = 1.0$). The third analysis for incongruent signal, a Bayesian analysis of 20 datasets constructed using a sliding window approach, recovered

Appendix 3: Publication version of Chapter 2

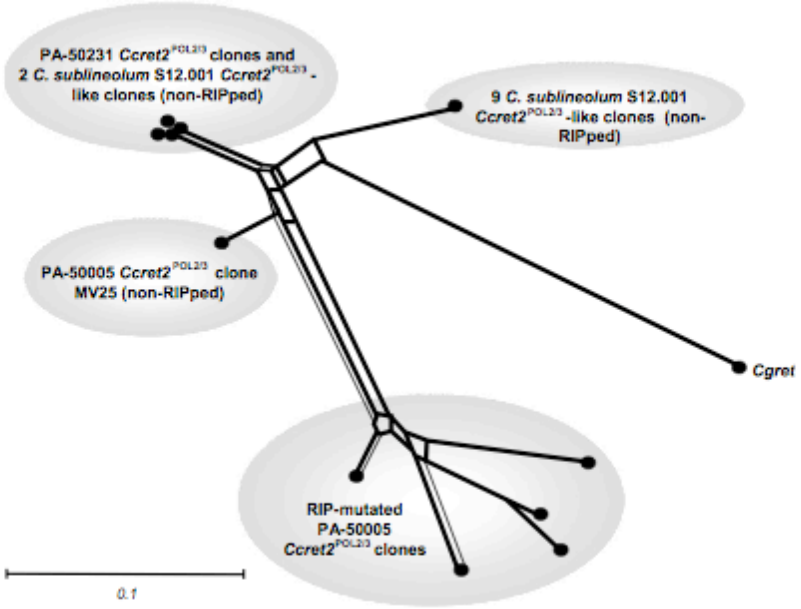


Fig. 7. *Ccret2*$^{POL2/3}$ network using LogDet distance analysis shows the pattern of reticulate relationships between the cloned sequences from *Colletotrichum cereale* isolates PA-50231 (clade A), PA-50005 (clade B), and *C. sublineolum* isolate S12001. The LTR-retrotransposon *Cgret* from *C. gloeosporioides* is included as an outgroup.

identical branching patterns among the *C. sublineolum* and *C. cereale* clade A elements across all 20 trees, while the clade B lineage was inconsistently structured from tree to tree (not shown). The likelihoods of the 20 trees, when used to perform an SH-likelihood ratio test, identified statistically significant levels of discordance between the topologies, indicative of either recombination or homoplasy (Table 3). Likewise, when the dataset was analyzed using the GARD multiple breakpoint method, a single recombination breakpoint was detected from the clade B sequences, while the remainder of the dataset presented no evidence for recombination.

Table 3
Results of the Shimodaira–Hasegawa likelihood ratio test of the 20 sliding window consensus trees from the *Ccret2*$^{Pol2/3}$ dataset used to statistically test whether discordant topologies were generated along the length of the nucleotide sequence alignment

Tree	−ln *L*	Diff. −ln *L*	*P*
1	2799.32	287.52	0.0000*
2	2513.44	1.65	0.9539
3	2511.79	(Best)	—
4	2512.59	0.79	0.9691
5	2607.40	95.61	0.1075
6	2635.69	123.89	0.0269*
7	2638.74	126.94	0.0240*
8	2760.59	248.80	0.0000*
9	3097.23	585.43	0.0000*
10	2730.17	218.37	0.0001*
11	2614.29	102.50	0.0688
12	2834.20	322.40	0.0000*
13	2647.30	135.50	0.0140*
14	2835.40	323.60	0.0000*
15	3130.95	619.15	0.0000*
16	2948.47	436.68	0.0000*
17	2795.08	283.29	0.0000*
18	2795.08	283.29	0.0000*
19	2809.97	298.17	0.0000*
20	2802.66	290.87	0.0000*

* $P < 0.05$.

4. Discussion

From an organismal standpoint, the lineage-specific distribution of RIP mutation in *C. cereale*—absent in clade A, present in clade B—is an important contribution to our understanding of how this species has evolved. The widespread identification of RIPping in diverse ascomycete species, including the closely related *C. falcatum* (J.A. Crouch and B.I. Hillman; unpublished data), strongly discourages the conclusion that the clade A genome might be "RIP-free", but our inability to detect a RIP signature clearly emphasizes the distinct nature of *C. cereale*'s main lineages and is consistent with previous observations about these groups. Using multilocus haplotype networks, Crouch et al. (2006) determined that *C. cereale* clade A is likely derived from a single ancestral haplotype. Such a population bottleneck, through the elimination of RIP-mutated

Appendix 3: Publication version of Chapter 2

alleles, could easily have produced the pattern of absentee RIP-mutated TEs observed in clade A, particularly if these sequences were of no selective advantage to the fungus. Furthermore, the many copies of RIP-mutated TEs in the clade B genome is entirely consistent with the reticulate population structure inferred for clade B using haplotype networks (Crouch et al., 2006), since both datasets suggest the same conclusion: that clade B has been influenced by gene flow processes. Gene flow as inferred through the haplotype network alone could have resulted from either sexual recombination or parasexually-derived mitotic recombination, but the detection of RIP-mutation in the clade B sequences—a process that occurs only during meiosis—provides compelling evidence of sexual recombination in the clade B lineage, where 14 of the 15 unique transposon sequences were RIPped. Moreover, the identification of at least five diverse *Ccret2*$^{POL2/3}$ groups structured along a distinct cline of RIP-mutation levels (Fig. 6) establishes that sexual recombination was not restricted to a single isolated event; instead, the sexual morph of the fungus must have been produced on multiple occasions, at least in the clade B lineage. But an important question leads to the next logical area of inquiry for this species: did the inferred clade B sexual recombination occur cryptically in extant populations or is the observed RIP signature merely the artifact of a long defunct sexual ancestral state?

For the genus *Colletotrichum* and its teleomorph *Glomerella*, the genetic basis of sexual compatibility is uncertain, despite the pioneering experiments of Edgerton, Wheeler and colleagues between 1914 and 1959 that served to elucidate the genetics of sexual development in *G. cingulata* (e.g. Driver and Wheeler, 1955; Edgerton, 1914; Edgerton et al., 1945; Lucas et al., 1944; Wheeler et al., 1948, 1959). What is well known, however, is that the regulation of mating in *Colletotrichum* is often unpredictable and appears to be rather different from that in other filamentous ascomycetes. While sexual recognition for most of the ascomycota is controlled by a one locus, two allele (idiomorph) mating system, generation of the *Glomerella* state is thought to be governed by at least two unlinked loci (e.g. *G. graminicola;* Vaillancourt et al., 2000) or by means of a single locus with multiple alleles (e.g. *G. cingulata*; Cisar and TeBeest, 1999). To date, extensive sampling of 11 *Colletotrichum* species, including *C. cereale*, has resulted in the detection of only a single idiomorph—the *MAT1-2*, regardless of whether the species in question is heterothallic or homothallic (Crouch et al., 2006; Chen et al., 2002; Du et al., 2005; Rodriguez-Guerra et al., 2005; J.A. Crouch and B.I. Hillman, unpublished data). Even the prediction of appropriate mating partners for genetic analysis in the genus can be difficult, since many species, including *G. graminicola*, *G. cingulata* and *G. lindemuthiana*, are both heterothallic and homothallic (Cisar et al., 1994; Cisar and TeBeest, 1999; Rodriguez-Guerra et al., 2005; Wheeler, 1954). Furthermore, individual isolates are frequently infertile, irrespective of mating conditions. Thus for *Colletotrichum*, the indirect assessment of recombination by means of molecular analysis, rather than through the direct experimental or natural observation of biological mating, is likely to yield a more accurate assessment of whether recombination has influenced populations of the fungus. The observation of RIP mutation in the present study suggests a greater level of complexity in populations of *C. cereale* than previously suspected, with the available evidence favoring the view that the evolution of *C. cereale* has been driven, at least in part, by recombination.

From a broad evolutionary perspective, the amino acid sequences of the transposons identified in this study—even those that are RIP-mutated—reflect phylogenetic relationships that are generally consistent with currently accepted taxonomy (Figs. 1–3). But as the analysis moved to the nucleotide level to examine the *Ccret2*$^{POL2/3}$ populations within individual genomes, inconsistencies began to emerge. The first irregularity, in which two of the 11 copies of *Ccret2*$^{POL2/3}$ sequenced from *C. sublineolum* were identical to copies of *Ccret2* from *C. cereale* clade A, might have resulted due to a retained ancestral polymorphism since it has been determined that *Ccret2* was present in the common ancestor of the sister species *C. cereale*, *C. sublineolum* and *C. falcatum* (J.A. Crouch, B.I. Hillman; unpublished data). Different copies of the TE might then have been randomly distributed during the subsequent speciation process. Horizontal gene transfer, which has been systematically established in only a few instances for eukaryotic TEs (Jordan et al., 1999; Daniels et al., 1990; Diao et al., 2006), provides an alternative explanation, and has been proposed to account for the unexpected presence of an active *Tad* element in *N. crassa* (Anderson et al., 2001).

A second irregularity appeared in the phylogenetic analysis of the *Ccret2*$^{POL2/3}$ dataset: the RIPped clade B *Ccret2*$^{POL2/3}$ elements clustered into five main groups, with each group presumably representing different numbers of RIP-mutation cycles acting on different copies of the element. The branches constructed from the RIPped TE copies defy cladistic assumptions, since these elements are established in the reverse chronological order of the actual evolutionary path within the tree topology. The most divergent, and in this case, also the oldest copies of *Ccret2*$^{POL2/3}$ are found at the tips of the tree, since the oldest copies of the element would have undergone repeated rounds of mutation relative to more recently inserted, less RIPped copies of the transposon (Fig. 6). In this scenario, although the basal taxa in the *Ccret2*$^{POL2/3}$ tree possess the more ancestral-like sequence, these transposon copies are actually the youngest examples of RIP-mutated elements within the sample.

The third anomaly in the *Ccret2*$^{POL2/3}$ dataset was concentrated within the RIPped clade B group, where numerous inconsistencies were observed. Several incongruent clade B topologies were recovered when the dataset was subjected to sliding window analysis, suggesting either the presence of recombination or RIP-induced homoplasy.

Appendix 3: Publication version of Chapter 2

Recombination was independently supported by the ϕ_w analysis, but since only a single recombination breakpoint was detected using the GARD method, it appears that the numerous incongruent topologies recovered using the sliding window analysis were the result of RIP mutation-derived substitution and rate heterogeneity rather than high levels of recombination. Our data cannot rule out the potential for recombination among the RIPped *Ceret2* elements, since recombination can be a powerful force acting on retrotransposons, as exemplified in *Saccharomyces* (for review, see Mieczkowski et al., 2006). The hypermutation of the RIPped sequences might even have diluted the signal of recombination in this dataset. Conversely, the accumulation of TEs in non-recombining regions of the genome is predicted both by population genetics theory (Montgomery et al., 1991; Langley et al., 1988; Charlesworth et al., 1994) and is also a common trend in many diverse organisms, including *Drosophila melanogaster*, *Tetraodon nigroviridis* and *Arabidopsis thaliana* (Fischer et al., 2004; Langley et al., 1988; Maside et al., 2005; Kapitonov and Jurka, 2003; Wright et al., 2003). In our dataset, it appears that RIP mutation-induced homoplasy rather than recombination is the primary driver of conflict.

In this study, the data clearly showed that the accelerated mutations experienced during the RIP process are not equivalent to the sequence variation experienced by their "normal" non-RIPped counterparts, with the rate of evolution experienced by different RIPped copies within a single *C. cereale* strain greatly exceeding the mutational forces experienced, even when compared to that which occurred on the intraspecific level between *C. cereale* and *C. sublineolum*. Phylogenetic analysis of the *Ceret2*$^{POL2/3}$ copies visibly illustrated the high levels of divergence of the RIPped TEs relative to non-mutated elements of the same mobile element (Fig. 6); the normal TE sequences showed little variability, but for the RIPped group, the tree topology was characterized by exceptionally long branches. Even if RIP is no longer an active defense in the genome of *C. cereale* clade B, once these elements are RIP-altered and inactivated, the selective pressures experienced by these genomic relics are likely to be no longer equivalent to those encountered by their non-mutated counterparts and the RIPped elements would evolve under an entirely different set of circumstances than their functional counterparts.

Comprehensive genome-scale studies of fungal transposon ecology and evolution have been largely confined to the yeasts, despite the demonstrated power of TEs to advance our understanding of the biology and evolution across a wide range of eukaryotes. Noteworthy discoveries that have greatly expanded our knowledge of genome variability and evolution have resulted from studies of TEs in plants, insects, mammals, and, in particular, humans (e.g. Brouha et al., 2003; Du et al., 2006; Nikaido et al., 2001; Roy-Engel et al., 2002; Vieira and Biemont, 2004; Vincent et al., 2003). RIP mutation will present a complicating factor in many fungal TE evolutionary analyses, as established both in the current study and in studies of *F. oxysporum* and *M. violaceum* (Hua-Van et al., 2001; Hood et al., 2005), but this is more than offset by the wealth of information gained. With the increasing availability of fungal genome sequence data (97 fungal genome sequencing projects are currently curated by the NCBI Entrez Genome Project), we have unprecedented opportunities to explore and interpret transposon distribution, function and diversity.

Acknowledgments

We thank Lisa Vaillancourt for providing the *C. graminicola* and *C. sublineolum* cultures used in this study. This work was funded by grants from the Rutgers Center for Turfgrass Science to B.I.H. and B.B.C and by the New Jersey Agricultural Experiment Station. We gratefully acknowledge financial support for J.A.C.'s graduate studies provided by a U.S. Environmental Protection Agency Science to Achieve Results (STAR) Graduate Fellowship, a Rutgers Excellence Fellowship, the Robert White-Stevens Fellowship and a Land Institute Natural Systems Agriculture Graduate Fellowship.

Although the research described in this article has been funded in part by the EPA's STAR fellowship program through Grant FP-91652101, it has not been subjected to any EPA review and therefore does not necessarily reflect the views of the Agency, and no official endorsement of any products or commercial services mentioned in this article should be inferred.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21, 2104–2105.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Anderson, C., Tand, Q., Kinsey, J., 2001. Elimination of active *Tad* elements during the sexual phase of *Neurospora crassa* life cycle. Fungal Genet. Biol. 33, 49–57.

Attard, A., Gout, L., Ross, S., Parlange, S., Cattolico, L., Balesdent, M.-H., Rouxel, T., 2005. Truncated and RIP-degenerated copies of the LTR retrotransposon Pholy are clustered in a pericentromeric region of the *Leptosphaeria maculans* genome. Fungal Genet. Biol. 42, 30–41.

Bender, J., Kleckner, N., 1992. Tn10 insertion specificity is strongly dependent upon sequences immediately adjacent to the target-site consensus sequences. Proc. Natl. Acad. Sci. USA 89, 7996–8000.

Bhat, A., Tamuli, R., Kasbekar, D., 2004. Genetic transformation of *Neurospora tetrasperma*, demonstration of repeat-induced point mutation (RIP) in self-crosses and a screen for recessive RIP-defective mutants. Genetics 167, 1155–1164.

Bouvet, G.F., Jacobi, V., Bernier, L., 2007. Characterization of three DNA transposons in the Dutch elm disease fungi and evidence of repeat-induced point (RIP) mutations. Fungal Genet. Biol. 44, 430–443.

Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V., Kazazian Jr., H.H., 2003. Hot L1s account for the bulk of retrotransposition in the human population. Proc. Natl. Acad. Sci. USA 100, 5280–5285.

Bruen, T.C., Phillipe, H., Bryant, D., 2006. A simple and robust statistical test to detect the presence of recombination. Genetics 172, 2665–2681.

Cambareri, E.B., Jensen, B.C., Schabtach, E., Selker, E.U., 1989. Repeat-induced G–C to A–T mutations in *Neurospora*. Science 244, 1571–1575.

Appendix 3: Publication version of Chapter 2

Cambareri, E.B., Singer, M.J., Selker, E.U., 1991. Recurrence of repeat-induced point mutation (RIP) in *Neurospora crassa*. Genetics 127, 699–710.

Chalker, D.L., Sandmeyer, S.B., 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. 6, 117–128.

Chalmers, R., Guhathakurta, A., Benjamin, H., Kleckner, N., 1998. IHF modulation of Tn10 transposition: sensory transduction of supercoiling status via a proposed protein/DNA molecular spring. Cell 93, 897–908.

Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371, 215–220.

Chen, F., Goodwin, P.H., Khan, A., Hsiang, T., 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris*. Can. J. Microbiol. 48, 427–436.

Cisar, C.R., Spiegel, F.W., TeBeest, D.O., Trout, C., 1994. Evidence for mating between isolates of *Colletotrichum gloeosporioides* with different host specificities. Curr. Genet. 25, 330–335.

Cisar, C.R., TeBeest, D.O., 1999. Mating system of the filamentous ascomycete, *Glomerella cingulata*. Curr. Genet. 35, 127–133.

Clutterbuck, A.J., 2004. MATE transposable elements in *Aspergillus nidulans*: evidence of repeat-induced point mutation. Fungal Genet. Biol. 41, 308–316.

Cogoni, C., Irelan, J.T., Schumacher, M., Schmidhauser, T.J., Selker, E.U., Macino, G., 1996. Transgene silencing of the *al-1* gene in vegetative cells of *Neurospora* is mediated by a cytoplasmic effector and does not depend on DNA–DNA methylation. EMBO J. 15, 3153–3163.

Crouch, J.A., Clarke, B.B., Hillman, B.I., 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. J. 10, 186–195.

Crouch, J.A., Clarke, B.B., Hillman, B.I., 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology 96, 46–60.

Daniels, S., Peterson, K., Strausbaugh, L., Kidwell, M., Chovnick, A., 1990. Evidence for horizontal transmission of the P transposable element between *Drosophila* species. Genetics 124, 339–355.

Dean, R.A., Talbot, N.J., Ebbole, D.J., Farman, M.L., Mitchell, T.K., Orbach, M.J., Thon, M., Kulkarni, R., Xu, J.-R., Pan, H., Read, N.D., Lee, Y.-H., Carbone, I., Brown, D., Oh, Y.Y., Donofrio, N., Jeong, J.S., Soanes, D.M., Djonovic, S., Kolomiets, E., Rehmeyer, C., Li, W., Harding, M., Kim, S., Lebrun, M.-H., Bohnert, H., Coughlan, S., Butler, J., Calvo, S., Ma, L.-J., Nicol, R., Purcell, S., Nusbaum, C., Galagan, J.E., Birren, B.W., 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature 434, 980–986.

Devine, S.E., Boeke, J.D., 1996. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev. 10, 620–633.

Diao, X., Freeling, M., Lisch, D., 2006. Horizontal transfer of a plant retrotransposon. PLoS Biol. 4, 0119–0128.

Driver, C., Wheeler, H., 1955. A sexual hormone in *Glomerella*. Mycologia 47, 311–316.

Du, C., Swigonova, Z., Messing, J., 2006. Retrotranspositions in orthologous regions of closely related grass species. BMC Evol. Biol. 6, 62.

Du, M., Schardl, C.L., Nuckles, E.M., Vaillancourt, L., 2005. Using mating-type gene sequences for improved phylogenetic resolution of *Colletotrichum* species complexes. Mycologia 97, 641–658.

Edgerton, C., 1914. Plus and minus strains in the genus *Glomerella*. Am. J. Bot. 1, 244–254.

Edgerton, C., Chilton, S., Lucas, G., 1945. Genetics of *Glomerella*. II. Fertilization between strains. Am. J. Bot. 32, 115–118.

Felsenstein, J., 2006. PHYLIP (Phylogeny Inference Package) version 3.6. Department of Genome Sciences, University of Washington, Seattle, Distributed by the author.

Fischer, C., Bouneau, L., Coutanceau, J., Weissenbach, J., Volff, J., Ozouf-Costaz, C., 2004. Global heterochromatic colocalization of transposable elements with minisatellites in the compact genome of the pufferfish *Tetraodon nigroviridis*. Gene 336, 175–183.

Galagan, J.E., Selker, E.U., 2004. RIP: the evolutionary cost of genome defense. Trends Genet. 20, 417–422.

Gardner, M., Hall, N., Fung, E., White, O., Berriman, M., Hyman, R., Carlton, J., Pain, A., Nelson, K., Bowman, S., Paulsen, I., James, K., Eisen, J., Rutherford, K., Salzberg, S., Craig, A., Kyes, S., Chan, M., Nene, V., Shallom, S., Suh, B., Peterson, J., Angiuoli, S., Pertea, M., Allen, J., Selengut, J., Haft, D., Mather, M., Vaidya, A., Martin, D., Fairlamb, A., Fraunholz, M., Roos, D., Ralph, S., McFadden, G., Cummings, L., Subramanian, G., Mungall, C., Venter, J., Carucci, D., Hoffman, S., Newbold, C., Davis, R., Fraser, C., Barrell, B., 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature 419, 498–511.

Goffeau, A., Barrell, B., Bussey, H., Davis, R., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J., Jacq, C., Johnston, M., Louis, E., Mewes, H., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S., 1996. Life with 6000 genes. Science 274, 563–567.

Graia, L., Lespinet, O., Rimbault, B., Dequard-Chablet, M., Choppin, E., Picard, M., 2001. Genome quality control: RIP (repeat-induced point mutation) comes to *Podospora*. Mol. Microbiol. 40, 586–595.

Hood, M.E., Katawczik, M., Giraud, T., 2005. Repeat-induced point mutation and the population structure of transposable elements in *Microbotryum violaceum*. Genetics 170, 1081–1089.

Hua-Van, A., Langin, T., Daboussi, M.-J., 2001. Evolutionary history of the *impala* transposon in *Fusarium oxysporum*. Mol. Biol. Evol. 18, 1959–1969.

Hua-Van, A., Le Rouzic, A., Maisonhaute, C., Capy, P., 2005. Abundance, distribution and dynamics of retrotransposons and transposons: similarities and differences. Cytogenet. Genome Res. 110, 426–440.

Huelsenbeck, J.P., Ronquist, F., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Huson, D.H., 1998. SplitsTree: analyzing and visualizing evolutionary data. Bioinformatics 14, 68–73.

Huson, D.H., Bryant, M.L., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267.

ICTV, 2004. Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses, Academic Press, New York.

International Human Genome Sequencing Consortium, 2001. Initial sequencing and analysis of the human genome. Nature 409, 860–921.

Ikeda, K.-I., Nakayashiki, H., Kataoka, T., Tamba, H., Hashimoto, Y., Tosa, Y., Mayama, S., 2002. Repeat-induced point mutation (RIP) in *Magnaporthe grisea*: implications for its sexual cycle in the natural field context. Mol. Microbiol. 45, 1355–1364.

International Mouse Genome Sequencing Consortium, 2001. Initial sequencing and comparative analysis of the mouse genome. Nature 420, 520–562.

Jordan, I.K., Matyunina, L.V., McDonald, J.F., 1999. Evidence for the recent horizontal transfer of long terminal repeat retrotransposon. Proc. Natl. Acad. Sci. USA 96, 12621–12625.

Kapitonov, V.V., Jurka, J., 2003. Molecular paleontology of transposable elements in the *Drosophila melanogaster* genome. Proc. Natl. Acad. Sci. USA 100, 6569–6574.

Ketting, R., Fisher, S., Plasterk, R., 1997. Target choice determinants of the Tc1 transposon of *Caenorhabditis elegans*. Nucleic Acids Res. 25, 4041–4047.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23, 1891–1901.

Langley, C., Montgomery, E., Hudson, R.R., Kaplan, N., Charlesworth, B., 1988. On the role of unequal exchange in the containment of transposable element copy number. Genet. Res. 52, 223–235.

Liu, G., Geurts, A., Yae, K., Srinivasan, A., Fahrenkrug, S., Largaespada, D., Takeda, J., Horie, K., Olson, W., Hackett, P., 2005. Target-site preferences of *Sleeping Beauty* transposons. J. Mol. Biol. 346, 161–173.

Lockhart, P.J., Steel, M.A., Hendy, M.D., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11, 605–612.

Appendix 3: Publication version of Chapter 2

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Bryant, S.H., 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Res. 33, D192–D196.

Lucas, G., Chilton, S., Edgerton, C., 1944. Genetics of *Glomerella*. I. Studies on the behavior of certain strains. Am. J. Bot. 31, 233–239.

Marchler-Bauer, A., Bryant, S.H., 2004. CD-Search: protein domain annotations on the fly. Nucleic Acids Res. 32, W327–W331.

Margolin, B.S., Garrett-Engele, P.W., Stevens, J.N., Fritz, D.Y., Garrett-Engele, C., Metzenberg, R.L., Selker, E.U., 1998. A methylated *Neurospora* 5S pseudogene contains a transposable element inactivated by repeat-induced point mutation. Genetics 149, 1787–1797.

Maside, X., Assimacopoulos, S., Charlesworth, B., 2005. Fixation of transposable elements in the *Drosophila melanogaster* genome. Genet. Res. 85, 195–203.

Messing, J., Dooner, H., 2006. Organization and variability of the maize genome. Curr. Opin. Plant Biol. 9, 157–163.

Mieczkowski, P., Lemoine, F., Petes, T., 2006. Recombination between retrotransposons as a source of chromosome rearrangements in the yeast *Saccharomyces cerevisiae*. DNA Repair 5, 1010–1020.

Montgomery, E., Huang, S., Langley, C., Judd, B., 1991. Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. Genetics 129, 1085–1098.

Montiel, M., Lee, H., Archer, D., 2006. Evidence of RIP (repeat-induced point mutation) in transposase sequences of *Aspergillus oryzae*. Fungal Genet. Biol. 43, 439–445.

Nakayashiki, H., Nishimoto, N., Ikeda, K., Tosa, Y., Mayama, S., 1999. Degenerate MAGGY elements in a subgroup of *Pyricularia grisea*: a possible example of successful capture of a genetic invader by a fungal genome. Mol. Gen. Genet. 261, 958–966.

Neuveglise, C., Sarfati, J., Latge, J.P., Paris, S., 1996. Afut1, a retrotransposon-like element from *Aspergillus fumigatus*. Nucleic Acids Res. 24, 1428–1434.

Nielson, M., Hermansen, T., Aleksenko, A., 2001. A family of DNA repeats in *Aspergillus nidulans* has assimilated degenerated retrotransposons. Mol. Gen. Genet. 265, 883–887.

Nikaido, M., Matsuno, F., Hamilton, H., Brownell, R.J., Cao, Y., Ding, W., Zuoyan, Z., Shedlock, A., Fordyce, R., Hasegawa, M., Okada, N., 2001. Retroposon analysis of major cetacean lineages: the monophyly of toothed whales and the paraphyly of river dolphins. Proc. Natl. Acad. Sci. USA 98, 7384–7389.

Paraskevis, D., Deforche, K., Lemey, P., Magiorkinis, G., Hatzakis, A., Vandamme, A.-M., 2005. SlidingBayes: exploring recombination using a sliding window approach based on Bayesian phylogenetic inference. Bioinformatics 21, 1274–1275.

Posada, D., Crandall, K.A., 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14, 817–818.

Pride, D. 2004. Swaap 1.0.1: a tool for analyzing substitutions and similarity in multiple alignments, distributed by the author.

Rodriguez-Guerra, R., Ramirez-Rueda, M., Cabral-Enciso, M., Garcia-Serrano, M., Lira-Maldonado, Z., Guevara-Gonzalez, R., Gonzalez-Chavira, M., Simpson, J., 2005. Heterothallic mating observed between Mexican isolates of *Glomerella lindemuthiana*. Mycologia 97, 793–803.

Roy-Engel, A.M., Carroll, M.L., El-Sawy, M., Salem, A.-H., Garber, R.K., Nguyen, S.V., Deininger, P.L., Batzer, M.A., 2002. Non-traditional Alu evolution and primate genomic diversity. J. Mol. Biol. 316, 1033–1040.

Sambrook, J., Russell, D., 2001. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, New York.

Selker, E.U., Cambareri, E.B., Jensen, B.C., Haack, K.R., 1987. Rearrangement of duplicated DNA in specialized cells of *Neurospora*. Cell 51, 741–752.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shiu, P.K., Raju, N.B., Zickler, D., Metzenberg, R.L., 2001. Meiotic silencing by unpaired DNA. Cell 107, 905–916.

Singleton, T.L., Levin, H.L., 2002. A long terminal repeat retrotransposon of fission yeast has strong preferences for specific sites of insertion. Eukaryot. Cell 1, 44–55.

Smiley, R.W., Dernoeden, P.H., Clarke, B.B., 2005. Compendium of Turfgrass Diseases. APS Press, St. Paul.

Swofford, D.L., 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. Nucleic Acids Res. 22, 4673–4680.

Thon, M., Pan, H., Diener, S., Papalas, J., Taro, A., Mitchell, T., Dean, R., 2006. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. Genome Biol. 7, R16.

Vaillancourt, L., Du, M., Wang, J., Rollins, J., Hanau, R., 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia 92, 430–435.

Vieira, C., Biemont, C., 2004. Transposable element dynamics in two sibling species: *Drosophila melanogaster* and *Drosophila simulans*. Genetica 120, 115–123.

Vincent, B.J., Myers, J.S., Ho, H.J., Kilroy, G.E., Walker, J.A., Watkins, W.S., Jorde, L.B., Batzer, M.A., 2003. Following the LINEs: an analysis of primate genomic variation at human-specific LINE-1 insertion sites. Mol. Biol. Evol. 20, 1338–1348.

Watters, M.K., Randall, T.A., Margolin, B.S., Selker, E.U., Stadler, D.R., 1999. Action of repeat-induced point mutation on both strands of a duplex and on tandem duplications of various sizes in *Neurospora*. Genetics 153, 705–714.

Wheeler, H., 1954. Genetics and evolution of heterothallism in *Glomerella*. Phytopathology 44, 342–345.

Wheeler, H., Driver, C., Campa, C., 1959. Cross- and self-fertilization in *Glomerella*. Am. J. Bot. 46, 361–365.

Wheeler, H., Olive, L., Ernest, C., Edgerton, C., 1948. Genetics of *Glomerella*: V. Crozier and ascus development. Am. J. Bot. 35, 722–728.

Wicker, T., Robertson, J., Schulze, S., Feltus, F., Magrini, V., Morrison, J., Mardis, E., Wilson, R., Peterson, D., Paterson, A., Ivarie, R., 2005. The repetitive landscape of the chicken genome. Genome Res. 15, 126–136.

Wright, S., Gagrawal, N., Bureau, T., 2003. Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. Genome Res. 13, 1897–1903.

Zhu, P., Oudemans, P.V., 2000. A long terminal repeat retrotransposon *Cg* ret from the phytopathogenic fungus *Colletotrichum gloeosporioides* on cranberry. Curr. Genet. 38, 241–247.

Zou, S., Ke, N., Kim, J.M., Voytas, D.F., 1996. The *Saccharomyces* retrotransposon Ty5 integrates preferentially into regions of silent chromatin at the telomeres and mating loci. Genes Dev. 10, 634–645.

Appendix 4: Publication version of Chapter 3

RESEARCH

Patterns of Diversity in Populations of the Turfgrass Pathogen *Colletotrichum cereale* as Revealed by Transposon Fingerprint Profiles

Jo Anne Crouch, Bernadette M. Glasheen, Wakar Uddin, Bruce B. Clarke, and Bradley I. Hillman*

ABSTRACT

Anthracnose disease of cool-season turfgrasses, caused by the fungus *Colletotrichum cereale*, has recently emerged as one of the most significant pathogens of *Poa annua*. Here we investigated the utility of four repetitive transposable elements as molecular markers for the analysis of *C. cereale* populations. Southern blot hybridization analysis revealed lineage-specific polymorphisms and distribution patterns for these transposons. Comparative phylogenetic analysis of three nonrepetitive protein coding DNA sequences against the transposon restriction fragment length polymorphisms indicated that the transposon sequences have similar evolutionary histories to those found in the sampled *C. cereale* population, despite the alteration of several transposon copies by repeat-induced point mutation. The variability and ubiquity of the *Ccret2*^{A15} transposon in *C. cereale* genomes suggest that this element could be used as a reliable DNA marker to discriminate between lineages of the fungus, identify hybrid genotypes, and analyze genetic diversity in populations of this turfgrass pathogen.

J.A. Crouch, B.M. Glasheen, B.B. Clarke, and B.I. Hillman, Dep. of Plant Biology and Pathology, Rutgers Univ., New Brunswick, NJ 08901-8520; W. Uddin, Dep. of Plant Pathology, The Pennsylvania State Univ., University Park, PA 16802. Received 2 Aug. 2007. *Corresponding author (hillman@aesop.rutgers.edu).

Abbreviations: ITS, intergenic transcribed spacer; kb, kilobase; PCR, polymerase chain reaction; RAPD, randomly amplified polymorphic DNA; RFLP, restriction fragment length polymorphism; RIP, repeat-induced point.

DURING THE PAST DECADE, the anamorphic fungus *Colletotrichum cereale* sensu lato Crouch, Clark and Hillman (formerly *C. graminicola* G.W. Wilson) (Crouch et al., 2006) emerged from relative obscurity to become one of the most devastating pathogens of the cool-season turfgrass *Poa annua*, causing epidemics of anthracnose disease in stands of this grasses maintained as golf course greens in North America (Smiley et al., 2005) and the United Kingdom (Mann and Newell, 2005). For golf course superintendents, management of anthracnose is a challenging and expensive undertaking. Control of the disease relies heavily on fungicide applications; however, resistance to benzimidazole, strobilurin, and sterol inhibitor fungicidal chemistries is an increasingly widespread phenomenon (Avila-Adame et al., 2003; Crouch et al., 2005; Wong and Midland, 2007; Wong et al., 2007; B.B. Clarke, unpublished data).

Because genetic variability between isolates of *C. cereale* may influence the trajectory of anthracnose disease of turfgrass, a comprehensive understanding of how *C. cereale* populations are organized and distributed across their geographic range could enhance the development and implementation of effective disease

Published in Crop Sci. 48:xxx–xxx (2008).
doi: 10.2135/cropsci2007.08.0427
© Crop Science Society of America
677 S. Segoe Rd., Madison, WI 53711 USA

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

Appendix 4: Publication version of Chapter 3

management strategies. At present only limited population-level data, derived from randomly amplified polymorphic DNA (RAPD) or isozyme markers, are available for the fungus (Backman et al., 1999; Browning et al., 1999; Chen et al., 2002; Horvath and Vargas, 2004), although two major lineages, designated clades A and B, have been recognized on the basis of intergenic transcribed spacer (ITS) nucleotide sequences (Crouch et al., 2005) and a multiple gene genealogical approach (Crouch et al., 2006). Currently, few apparent biological patterns are readily ascribable to this divergence, and uncertainty exists as to whether the two groups are genetically isolated. *Colletotrichum cereale* clades A and B are morphologically indistinguishable and have overlapping distributions; furthermore, each lineage includes a cohort of both disease-inducing isolates from turfgrass species and their nonpathogenic counterparts from cereal crops and natural grassland ecosystems (Crouch et al., 2006; J.A. Crouch and B.I. Hillman, unpublished data).

The presence or absence of transposons at particular loci is a major contributor to restriction fragment length polymorphism (RFLP) variation in filamentous fungi. The primary objective of this research was to determine if repetitive transposable elements from the *C. cereale* genome could be developed as molecular markers to assess population structure and variability in the species. In the present study, we evaluated four elements representing three species of transposons (Crouch et al., 2007) from *C. cereale* as molecular markers to examine population structure in this organism. Because of their ubiquitous and repetitive nature, molecular marker systems based on mobile transposable element polymorphisms have been used for population-level analyses of numerous organisms, including several filamentous fungi (Diez et al., 2003; Farman et al., 1996; Girard and Freeling, 1999; Kohn et al., 1991; Linder-Basso et al., 2001; Milgroom et al., 1992). The presence of a transposon at a genomic locus is typically a good indicator of identity by descent, while the absence of an element at a site is recognized as the ancestral state. Transposon insertional RFLP data can be relatively free of homoplasic data that might be inconsistent with an organism's evolutionary history, since the independent insertion of two different transposon copies at the exact same location on a chromosome is extremely unlikely. The parallel loss of transposon copies through excision or homologous recombination may be problematic, however (Carbone et al., 1999), and alteration by repeat-induced point (RIP) mutation of transposons may complicate the evolutionary signal (Crouch et al., 2007). Although base substitutions in the restriction enzyme recognition sequence can theoretically generate nonhomologous bands of identical size, there is little likelihood of this type of convergence in groups of closely related individuals where evolutionary rates are low; in particular, parallel independent gains of restriction sites occur with only a small probability (Nei and Tajima, 1983; Nei and Tajima, 1985; Upholt, 1977).

The objectives of this study were to determine if transposon RFLP markers support the separation of *C. cereale* into two distinct lineages as previously described (Crouch et al., 2005, 2006) and to examine whether these markers offer any advantages over nucleotide sequence data in discerning structure in *C. cereale* populations. In particular, we considered to what extent these transposons could extend our understanding of how the major *C. cereale* lineages have evolved.

MATERIALS AND METHODS

Fungal Cultures

Twenty-one single spore cultures of *C. cereale* were isolated from diseased *Poa annua* on 11 golf course greens located within a 100-km radius in Pennsylvania (Fig. 1, Table 1) and cultured as previously described (Crouch et al., 2006). Isolates of *C. graminicola* from *Zea mays*, *C. sublineolum* from *Sorghum bicolor*, and *C. falcatum* from *Saccharum officinarum* were used for outgroup comparisons.

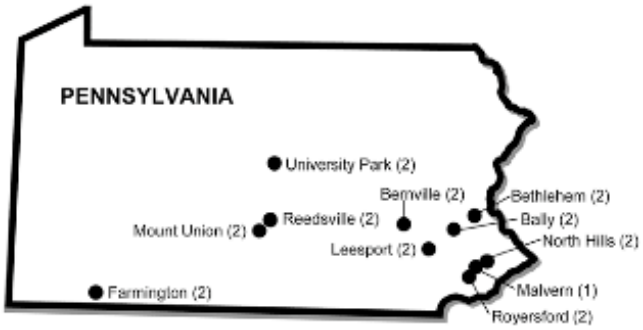


Figure 1. Map of Pennsylvania, illustrating the origination of the *Colletotrichum cereale* isolates used in this study. The number of isolates from each location is listed in parentheses after the location name.

RFLP Analyses

Genomic DNA was isolated from mycelia using a standard phenol:chloroform extraction protocol (Sambrook and Russell, 2001). *Hin*dIII-digested genomic DNA was size fractionated by gel electrophoresis for 18 h at 45v in 1x TBE buffer, then visualized using ethidium bromide staining. Southern blots for RFLP analysis were prepared by transferring the DNA to Zeta-Probe membranes (Bio-Rad, Hercules, CA) using a Posiblot Pressure Blotter (Strategene, La Jolla, CA) at 75 mm Hg. Five hundred nanograms of polymerase chain reaction (PCR) amplicon from each of the four individual transposon sequences (Table 2) were radiolabeled with [α^{32}P]dCTP (MP Biomedicals, Irvine, CA) using the Random Primers DNA Labeling System (Invitrogen Corp., Carlsbad, CA). Hybridizations were performed as previously described (Crouch et

Appendix 4: Publication version of Chapter 3

Table 1. Fungal strains used in this study. All fungi were isolated from *Poa annua* unless otherwise noted.

Isolate name	Fungal species	Location	*Colletotrichum cereale* clade	*Collect1* DNA transposon I29 sequence†	*Ccret2 Metaviridae* retrotransposon DBP16 sequence*	*Ccret1 Pseudoviridae* retrotransposon DBP6 sequence*	*Ccret2 Metaviridae* retrotransposon A15 sequence
PA-V1	*C. cereale*	University Park, PA	A	–	–	–	+
PA-V2	*C. cereale*	University Park, PA	A	–	–	–	+
PA-WH3	*C. cereale*	Leesport, PA	A	–	–	–	+
PA-WH4	*C. cereale*	Leesport, PA	A	–	–	–	+
PA-50111	*C. cereale*	Royersford, PA	A	–	–	–	+
PA-50114	*C. cereale*	Royersford, PA	A	–	–	–	+
PA-50014	*C. cereale*	Malvern, PA	A	–	–	–	+
PA-50231	*C. cereale*	North Hills, PA	A	–	–	–	+
PA-50234	*C. cereale*	North Hills, PA	A	–	–	–	+
PA-50101	*C. cereale*	Mount Union, PA	A	–	–	–	+
PA-50103	*C. cereale*	Mount Union, PA	A	–	–	–	+
PA-50183	*C. cereale*	Reedsville, PA	A	+	+	+	+
PA-50181	*C. cereale*	Reedsville, PA	B	+	+	+	+
PA-B211	*C. cereale*	Bally, PA	A	–	–	–	+
PA-B4410	*C. cereale*	Bally, PA	B	+	+	+	+
PA-50002	*C. cereale*	Bernville, PA	B	+	+	+	+
PA-50005	*C. cereale*	Bernville, PA	B	+	+	+	+
PA-50621	*C. cereale*	Farmington, PA	B	+	+	+	+
PA-50623	*C. cereale*	Farmington, PA	B	+	+	+	+
PA-S1112	*C. cereale*	Bethlehem, PA	B	+	+	+	+
PA-S2113	*C. cereale*	Bethlehem, PA	B	+	+	+	+
IN-900190	*C. graminicola*	Indiana	–	–	–	–	–
MO-100178	*C. graminicola*	Missouri	–	–	–	–	–
S.3001	*C. sublineolum*	*Burkina Fasso*	–	–	–	–	+
S.12001	*C. sublineolum*	Brazil *(Sorghum bicolor)*	–	–	–	–	+
MAFF306170	*C. falcatum*	Japan *(Saccharum officinarum)*	–	–	+	–	+
MAFF306299	*C. falcatum*	Japan *(Saccharum officinarum)*	–	–	+	–	+
MAFF305077	*C. falcatum*	Japan *(Saccharum officinarum)*	–	–	–	–	+

†Repeat-induced point mutated sequence.

al., 2007). Hybridized membranes were exposed to autoradiography film (Lab Scientific, Inc., Livingston, NJ) in the presence of an intensifying screen and incubated at −70°C for 24 to 48 h before development. We evaluated RFLP banding patterns of four sequences from three transposon species (*Collect1*^{I29}, *Ccret-1*^{DBP6}, *Ccret2*^{DBP16}, and *Ccret2*^{A15}).

Because the retrotransposon sequences *Ccret1*^{DBP6} and *Ccret2*^{A15} were identified in all of the *C. cereale* isolates sampled, the RFLP patterns from these elements were used to discern patterns of population subdivision. Bands on the autoradiograms were scored visually as either present or absent and coded as binary data. The datasets were analyzed to identify population groupings using the Bayesian Monte Carlo Markov chain-based clustering program Structure 2.1 (Falush et al., 2003; Pritchard et al., 2000) for 1,000,000 repetitions each, with the first 20,000 discarded as burn-in. These analyses were run using the admixture model and correlated allele frequencies between populations, which is considered the best strategy for detecting subtle differences in population structure (Falush et al., 2003). The degree of α admixture was empirically derived from the data, and the distribution of allelic frequencies λ was set to 1 (Falush et al., 2003). Twenty runs were performed for $K = 1$ through 10 (where K = the maximum number of populations).

Phylogenetic Analyses

Phylogenetic analysis was performed using three nuclear loci previously shown capable of differentiating between the two major lineages of *C. cereale*, with PCR amplified fragments of the *ITS1/5.8S/ITS2* ribosomal DNA (ITS), the HMG-box of the *Mat-1-2* mating idiomorph (HMG), and the manganese superoxide dismutase (*Sod2*) genes used to generate nucleotide sequence data as previously described (Crouch et al., 2006). The sister species of *C. cereale*—*C. sublineolum* and *C. falcatum*—along with the more distantly related species, *C. graminicola* (Crouch et al., 2006; J.A. Crouch and B.I. Hillman, unpublished data) were included as outgroup taxa. Multiple sequence alignments were constructed using Clustal W (Thompson et

Appendix 4: Publication version of Chapter 3

Table 2. Primer sequences used in this study.

Gene	Clone	Lineages amplified	Primer Name	Sequence
rDNA: ITS1, 5.8S, ITS2	–	A, B	ITS 4	White et al. (1990
rDNA: ITS1, 5.8S, ITS2	–	A, B	ITS 5	White et al. (1990)
Mat 1–2: HMG box	–	A, B	cgHMG-F	Vaillancourt et al. (2000)
Mat 1–2: HMG box	–	A, B	cgHMG-R	Vaillancourt et al. (2000)
Sod-2	–	A, B	SOD625-F	Crouch et al. (2006)
Sod-2	–	A, B	SOD625-R	Crouch et al. (2006)
Ccret2 LTR retrotransposon	A15	A, B	MV-POL-2F	Crouch et al. (2007)
Ccret2 LTR retrotransposon	A15	A, B	MV-POL-3R	Crouch et al. (2007)
Ccret2 LTR retrotransposon	DBP16	B	MV-GAG-20F	Crouch et al. (2007)
Ccret2 LTR retrotransposon	DBP16	B	MV-GAG-21R	Crouch et al. (2007)
Ccret1 LTR retrotransposon	DBP6	B	PV-INT-40F	Crouch et al. (2007)
Ccret1 LTR retrotransposon	DBP6	B	PV-INT-41R	Crouch et al. (2007)
Collect1 DNA transposon	I-29	B	*pogo*-20F	Crouch et al. (2007)
Collect1 DNA transposon	I-29	B	*pogo*-21R	Crouch et al. (2007)

al., 1994) as launched in MegAlign (DNASTAR, Inc., Madison, WI), and manually adjusted to exclude gaps and ambiguously aligned regions from the dataset. Tree topologies were estimated from the combined multilocus nucleotide sequence dataset in MrBayes v.3.1 (Huelsenbeck and Ronquist, 2003) by performing two simultaneous runs of one cold and three heated Metropolis-coupled Monte Carlo Markov chains (MCMCMC) for 40,000,000 generations and sampling trees every 500 generations. Each individual gene region was partitioned in the analysis, and a general evolutionary model for each partition was incorporated as selected using the program ModelTest v.3.06 (Posada and Crandall, 1998) (ITS model: TrNef+G, A→G 1.5282, C→T 3.9607; $\alpha = 0.1317$; HMG model: HKY+G, A = 0.2654, C = 0.2953, G = 0.2622, T = 0.1770; Ti/Tv = 1.2783; $\alpha = 1.50421$; *Sod2* model: TrN+I, A = 0.2477, C = 0.3035, G = 0.2612, T = 0.1876; A→G 5.0186, C→T 4.8842, G→T 1.0; Pinv = 0.4998; equal rates for all sites). Run diagnostics were performed every 1000th generation, with the first 25% of the output discarded as burn-in. To confirm that the two separate runs converged to a stationary distribution, the average standard deviation of split frequencies, the potential scale reduction factor convergence diagnostic, and the plot of generation versus log likelihood were each examined for values and distributions characteristic of samples drawn from a posterior probability distribution. Trees sampled from the posterior distribution were imported into PAUP* v.4.0b10 (Swofford, 2000) and used to construct 75% majority-rule consensus trees.

Nucleotide Sequences

All new sequences generated by this study have been deposited in the National Center for Biotechnology Information GenBank database (accession numbers DQ663514–DQ663534).

RESULTS

Phylogenetic Assessment of Populations Using Sequence Data

Although drawn from a geographically limited region in Pennsylvania, the fungal specimens included in this study represent both of the major *C. cereale* evolutionary lineages, clades A and B, allowing us to test whether the multilocus RFLP banding patterns of four sequences from three transposon species (*Collect1*^{I29}, *Ccret1*^{DBP6}, *Ccret2*^{DBP16}, and *Ccret2*^{A15}) could be used to distinguish the major lineages in this species, even on a relatively fine scale. Three of the probes—*Collect1*^{I29}, *Ccret1*^{DBP6}, *Ccret2*^{DBP16}—have been altered in the past through RIP mutation, a genome defense system deployed by filamentous fungi that produces C→T and G →A transitions in repetitive DNA (Cambareri et al., 1989), including transposable elements. To evaluate the transposon-based population hypotheses, a strict consensus tree of 21 *C. cereale* isolates was constructed from 33,206 trees using Bayesian estimates from the combined ITS/HMG/*Sod2* dataset (Fig. 2). Both lineages were represented in the tree topology and supported by posterior probabilities of 100, with 13 isolates from *C. cereale* clade A and 8 isolates from clade B. Two of the geographic locations contained isolates from each of the two clades.

Limited Distribution of the TE Sequences *Collect1*^{I29} and *Ccret2*^{DBP16}

The TE markers *Collect1*^{I29} and *Ccret2*^{DBP16} produced fingerprint profiles largely consistent with the phylogenetic groups and confirmed the repetitive nature of the transposon sequences when hybridized against the restricted DNA gel blots. Isolates phylogenetically characterized as *C. cereale* clade B resulted in ~25 hybridizing bands on the autoradiograms (Fig. 3) with little polymorphism observed between the individual isolates. In contrast, all clade A isolates except PA-50183 were devoid of the *Collect1*^{I29} and *Ccret2*^{DBP16} sequences, as were the outgroup samples of *C. graminicola* and *C. sublineolum*. Polymerase chain reaction amplification using several alternate primer pairs from *Collect1*^{I29} and *Ccret2*^{DBP16} recovered the same pattern of presence or absence, failing to yield a product in clade A isolates even under conditions of low stringency (data not shown). The presence of these two elements in the genomes of *C. cereale* clade B and not in clade A is consistent with the fact that both of these transposon sequences are extensively RIP mutated, a process that has not been observed for clade A strains of the fungus (Crouch et al., 2007). But the PCR-based identification of *Ccret2*^{DBP16} from two of the three *C. falcatum* outgroup strains (data not shown) suggests that this RIP-mutated element was already present in the common ancestor of *C. falcatum* and *C. cereale* and was subsequently lost from *C. cereale* clade A after its divergence from clade B (Fig. 4).

Appendix 4: Publication version of Chapter 3

The Retrotransposons *Ccret1*^{DBP6} and *Ccret2*^{A15} Are Found in Both *C. cereale* Lineages

In contrast to the limited distribution of *Collect1*^{I29} and *Ccret2*^{DBP16} within the species, Southern blot analysis of the *C. cereale* population (Fig. 3) using the RIP-mutated *Ccret1*^{DBP6} probe revealed the presence of this retrotransposon in both of the major *C. cereale* lineages, although PCR amplification using a range of high and low stringency conditions and primer pairs demonstrated that *Ccret1*^{DBP6} was absent from the DNA of *C. graminicola*, *C. sublineolum*, and *C. falcatum* (data not shown). Each of the *C. cereale* clades exhibited visually distinct banding patterns. Clade B isolates yielded between 9 and 15 *Ccret1*^{DBP6} bands ranging in size from ~0.5 to 9 kilobase (kb), but with the exception of isolate PA-50183, the clade A isolates faintly hybridized at only one or two restriction fragments. Low copy number of *Ccret1*^{DBP6} in the genome of clade A isolates was anticipated since analysis of the element from a cosmid library found that this retrotransposon is present only as two unmutated copies at a single genomic locus in clade A isolate NJ-6340 (Crouch et al., 2007). The observed faint hybridization to the RIP-mutated probe sequence was similarly predicted from the cosmid sequence data since this transposon was not found to be RIP-mutated in clade A (Crouch et al., 2007). All *C. cereale* isolates shared the 1-kb *Ccret1*^{DBP6} band, indicating that this is probably the ancestral locus of *Ccret1*^{DBP6} and that subsequent amplification and RIP-mutation of this retrotransposon occurred only after the divergence of clades A and B (Fig. 4).

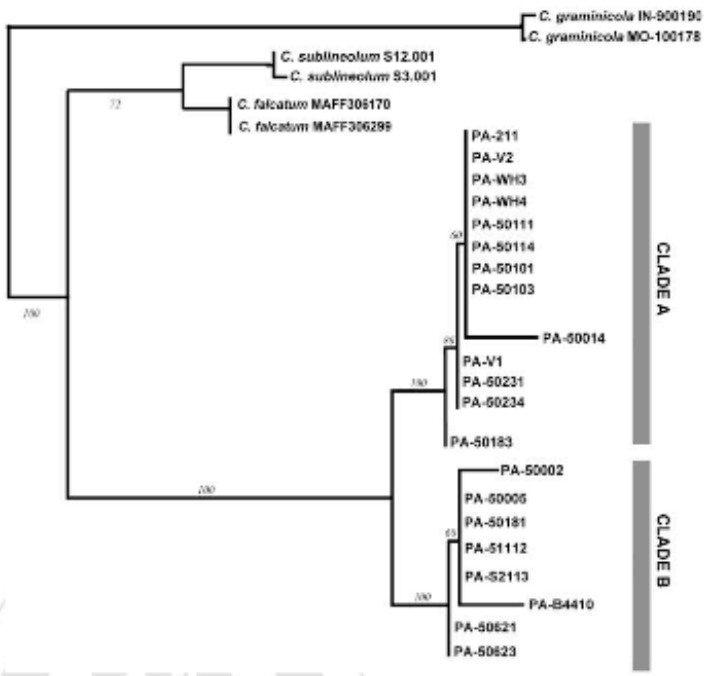


Figure 2. Multilocus tree estimated through Bayesian phylogenetic analysis of three protein coding genes supporting the division of the *Colletotrichum cereale* isolates into two main lineages, clades A and B (–*ln* likelihood = 3430.89).

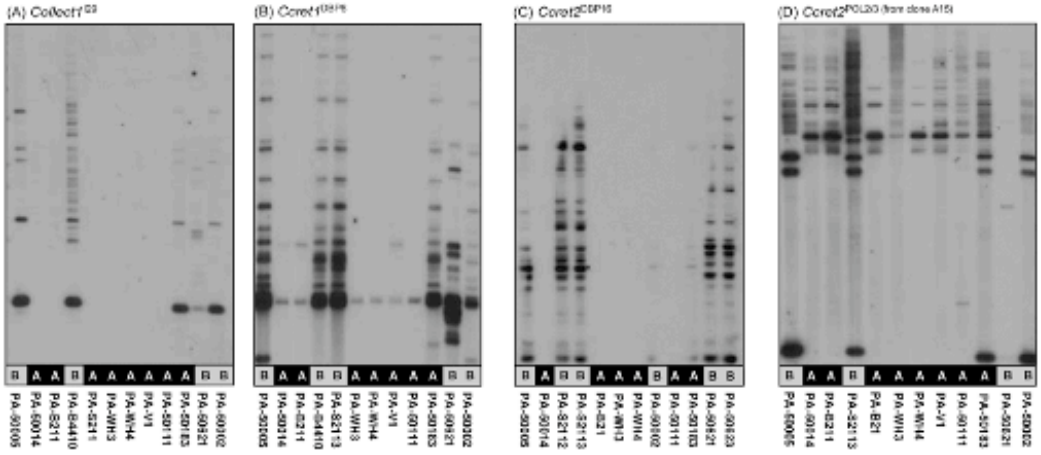


Figure 3. Southern blot hybridizations of *Hin*dIII-digested genomic DNA from a representative sample of *Colletotrichum cereale* clade A and B isolates using four transposon sequences as the probe. (A) *Collect1*^{I29} DNA transposon, (B) *Ccret1*^{DBP6} retrotransposon, (C) *Ccret2*^{DBP16} retrotransposon, (D) *Ccret2*$^{POL2/3}$ (from clone A15) retrotransposon.

Appendix 4: Publication version of Chapter 3

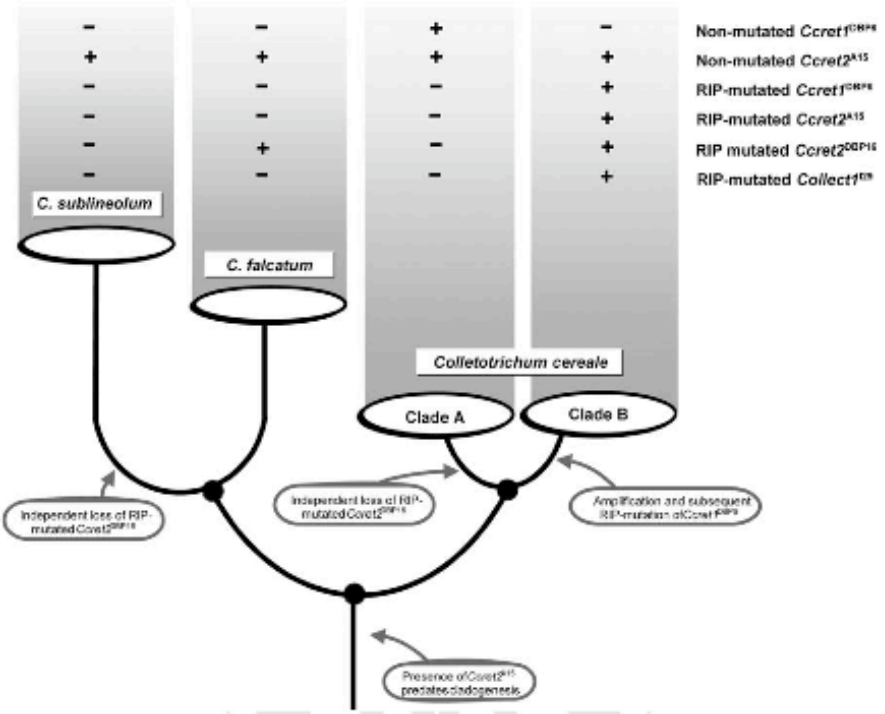


Figure 4. A schematic tree showing the presence or absence of the transposons evaluated in this study. (RIP = repeat-induced point.)

The *Ccret2*^{A15} retrotransposon sequence was the only transposon used as a probe in this study that was not RIP-altered, although in clade B strains of the fungus, this element can be present as both RIPped and non-RIPped variants within a single genome (Crouch et al., 2007). Of the four sequences evaluated, *Ccret2*^{A15} was the only transposon that produced a polymorphic RFLP banding pattern (Fig. 3). Like the other three transposon probes, the *Ccret2*^{A15} marker produced a visually distinctive banding pattern clearly differentiating between isolates belonging to phylogenetic clades A and B. Likewise, clade A isolate PA-50183 exhibited the clade B-like fingerprint rather than the clade A-like pattern predicted by phylogenetic affiliation. Polymerase chain reaction amplification identified *Ccret2*^{A15} from one of the two *C. sublineolum* isolates and all three of the *C. falcatum* isolates; however, it was absent from the more distantly related *C. graminicola*, suggesting that this transposon sequence was present in the common ancestor of *C. cereale*, *C. sublineolum*, and *C. falcatum* (Fig. 4).

Estimates of Population Subdivision Using the Retrotransposon RFLP Datasets

Since the *Ccret1*^{DBP6} and *Ccret2*^{A15} sequences were present in all of the *C. cereale* isolates sampled for this study, binary datasets were generated by coding the banding patterns produced by these elements as either present or absent to evaluate population subdivision. We first used the binary datasets to determine if the retrotransposon distribution within the genome was congruent with the HMG/ITS/*Sod2* evolutionary hypothesis. Consistent with the phylogenetic tree topology and the visual estimations made from the autoradiograms, two distinct populations, corresponding to clades A and B, were inferred from the RFLP datasets using the Bayesian clustering method implemented in the program Structure (Pritchard et al., 2000).

DISCUSSION

Consistent with the multilocus phylogenetic tree topology (Fig. 2), all four transposon RFLP fingerprint patterns recovered the division of *C. cereale* into two main lineages as previously established for the species (Crouch et al., 2006), either through distinct banding patterns or by their presence or absence. The only inconsistency observed between the nucleotide sequence data set and the transposon RFLPs was the manifestation of clade B-like banding patterns for the clade A isolate PA-50183 by all four transposon markers (Fig. 4), suggesting that this isolate may be a hybrid between the two lineages. Despite the potential for RIP-induced homoplasy in these analyses (Crouch et al., 2007), our data showed the *C. cereale* transposon RFLP signal in

Appendix 4: Publication version of Chapter 3

these analyses to be largely congruent with the non-TE datasets, with both the RFLPs and sequence analysis of three protein-coding genes yielding the same general conclusions. Although none of the RFLP fingerprints predicted any further population substructure beyond the two main lineages, this is likely a reflection of the small, geographically limited sample size evaluated in this study rather than a lack of sensitivity on the part of the markers. Since the purpose of this study was to determine whether transposon RFLP patterns are suitable molecular markers rather than drawing conclusions about the genetic makeup of populations, further study will be required to make this determination.

The interspecific distribution and intraspecific polymorphic banding patterns demonstrated that of the four markers evaluated, *Coret2*^{A15} sequence has the potential to serve as an effective RFLP marker for future population analysis of *C. cereale* and may even be adopted for use in populations of the closely related, economically important plant pathogens *C. sublineolum* and *C. falcatum*. *Coret2*^{A15} is polymorphic and was present in all *C. cereale* isolates sampled in this study; additionally, PCR-based screening shows that this transposon is widely distributed across the geographic range for this species and is present in both turfgrass pathogenic strains as well as *C. cereale* isolated from prairie, forage, and cereal crops (J.A. Crouch and B.I. Hillman, unpublished data). In contrast, while any of the other three transposons surveyed in this work—*Collect1*^{I29}, *Coret1*^{DBP6} and *Coret2*^{DBP16}—might in theory be used to evaluate populations of *C. cereale* clade B given the polymorphic banding patterns shown by the group, the high level of RIP mutation that characterizes these elements renders the use of these transposons as RFLP markers potentially problematic (Crouch et al., 2007). Under normal circumstances, although base substitutions in a restriction enzyme recognition sequence can theoretically generate nonhomologous bands of identical size, there is little likelihood of this type of convergence in groups of closely related individuals where evolutionary rates are low; in particular, parallel independent gains of restriction sites with six-base recognition sequences are have been found to occur with only a small probability (Nei and Tajima, 1983, 1985; Upholt, 1977). But for RIPped transposons, restriction sites are more rapidly gained or lost since overall nucleotide composition and dinucleotide patterns are skewed, often occurring at a range of different levels contingent on how many rounds of RIP mutation have acted on a given element. Thus, because RIP mutation has been found to act on these transposons, we cannot exclude the possibility that the different allelic states (±) observed at each locus are merely artifacts of RIP alterations rather than accurately reflecting common descent. For these reasons, for *C. cereale* clade B and other fungi where there is evidence of RIP mutation, transposon RFLP datasets should be regarded as potentially homoplasic unless independently derived support exists for the interpretation of homology. In the present study, however, the agreement between transposon RFLP data and the three independent protein coding genes attest to the consistency of the RFLP data in this sampled population and suggest that the *Coret2*^{A15} transposon-based marker can serve as a valuable tool in future population studies of *C. cereale*.

Acknowledgments

We thank Lisa Vaillancourt for providing the *Colletotrichum graminicola* and *C. sublineolum* cultures and the National Institute of Agrobiological Sciences Genebank of Ibaraki, Japan, for the *C. falcatum* cultures used in this study. This work was funded by grants from the Rutgers Center for Turfgrass Science to B.I.H. and B.B.C. and by the New Jersey Agricultural Experiment Station. We gratefully acknowledge financial support for J.A.C.'s graduate studies provided by a U.S. Environmental Protection Agency Science to Achieve Results (STAR) Graduate Fellowship, the Ralph Geiger Endowment, a Rutgers Excellence Fellowship, the Robert White-Stevens Fellowship, the Peter Selmer Loft Memorial Scholarship fund, and a Land Institute Natural Systems Agriculture Graduate Fellowship. Although the research described in this article has been funded in part by the USEPA's STAR fellowship program through grant FP-91652101, it has not been subjected to USEPA review and therefore does not necessarily reflect the views of the agency, and no official endorsement of any products or commercial services mentioned in this article should be inferred.

References

Avila-Adame, C., G. Olaya, and W. Koller. 2003. Characterization of *Colletotrichum graminicola* isolates resistant to strobulurin-related QoI fungicides. Plant Dis. 87:1426–1432.

Backman, P.A., P.J. Landschoot, and D.R. Huff. 1999. Variation in pathogenicity, morphology, and RAPD marker profiles in *Colletotrichum graminicola* from turfgrasses. Crop Sci. 39:1129–1135.

Browning, M., L.V. Rowley, P. Zeng, J.M. Chandlee, and N. Jackson. 1999. Morphological, pathogenic, and genetic comparisons of *Colletotrichum graminicola* isolates from *Poaceae*. Plant Dis. 83:286–292.

Cambareri, E.B., B.C. Jensen, E. Schabtach, and E.U. Selker. 1989. Repeat-induced G-C to A-T mutations in *Neurospora*. Science 244:1571–1575.

Carbone, I., J. Anderson, and L. Kohn. 1999. Patterns of descent in clonal lineages and their multilocus fingerprints are resolved with combined gene genealogies. Evolution 53:11–21.

Chen, F., P.H. Goodwin, A. Khan, and T. Hsiang. 2002. Population structure and mating-type genes of *Colletotrichum graminicola* from *Agrostis palustris*. Can. J. Microbiol. 48:427–436.

Crouch, J.A., B.B. Clarke, and B.I. Hillman. 2005. Phylogenetic relationships and fungicide sensitivities of *Colletotrichum graminicola* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. J. 10:186–195.

Crouch, J.A., B.B. Clarke, and B.I. Hillman. 2006. Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and corn. Phytopathology 96:46–60.

Crouch, J.A., B.M. Glasheen, M.A. Giunta, B.B. Clarke, and B.I. Hillman. 2007. The evolution of transposon repeat-induced

Appendix 4: Publication version of Chapter 3

point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. Fungal Genet. Biol. (in press), doi:10.1016/jfgb.2007.08.004.

Diez, J., T. Beguiristain, F. Le Tacon, J.M. Casacuberta, and D. Tagu. 2003. Identification of Ty1-copia retrotransposons in three ectomycorrhizal basidiomycetes: Evolutionary relationships and their use as molecular markers. Curr. Genet. 43:34–44.

Falush, D., M. Stephens, and J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Farman, M.L., S. Taura, and S.A. Leong. 1996. The *Magnaporthe grisea* DNA fingerprinting probe MGR586 contains the 3′ end of an inverted repeat transposon. Mol. Gen. Genet. 251:675–681.

Girard, L., and M. Freeling. 1999. Regulation changes as a consequence of transposon insertion. Dev. Genet. 25:291–296.

Horvath, B.J., and J.M. Vargas. 2004. Genetic variation among *Colletotrichum graminicola* isolates from four hosts using isozyme analysis. Plant Dis. 88:402–406.

Huelsenbeck, J.P., and F. Ronquist. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Kohn, L.M., E. Stasovski, I. Carbone, J. Royer, and J.B. Anderson. 1991. Mycelial incompatibility and molecular markers identify genetic variability in field populations of *Sclerotinia sclerotiorum*. Phytopathology 81:480–485.

Linder-Basso, D., R. Foglia, P. Zhu, and B.I. Hillman. 2001. *Crypt1*, an active *Ac*-like transposon from the chestnut blight fungus, *Cryphonectria parasitica*. Mol. Gen. Genet. 265:730–738.

Mann, R.L., and A.J. Newell. 2005. A survey to determine the incidence and severity of pests and diseases on golf course putting greens in England, Ireland, Scotland, and Wales. Int. Turfgrass Soc. Res. J. 10:224–229.

Milgroom, M.G., S.E. Lipari, and W.A. Powell. 1992. DNA fingerprinting and analysis of population structures of the chestnut blight fungus, *Cryphonectria parasitica*. Genetics 131:297–306.

Nei, M., and F. Tajima. 1983. Maximum likelihood estimation of the number of nucleotide substitutions from restriction site data. Genetics 105:207–217.

Nei, M., and F. Tajima. 1985. Evolutionary change of restriction cleavage sites and phylogenetic inference for man and apes. Mol. Biol. Evol. 2:189–205.

Posada, D., and K.A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. Bioinformatics 14:817–818.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Sambrook, J., and D. Russell. 2001. Molecular cloning: A laboratory manual. 3rd ed. Cold Spring Harbor Laboratory, New York.

Smiley, R.W., P.H. Dernoeden, and B.B. Clarke. 2005. Compendium of turfgrass diseases. 3rd ed. APS Press, St. Paul, MN.

Swofford, D.L. 2000. PAUP*: Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, MA.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and matrix choice. Nucleic Acids Res. 22:4673–4680.

Upholt, W.B. 1977. Estimation of DNA sequence divergence from comparison of restriction endonuclease digests. Nucleic Acids Res. 4:1257–1265.

Vaillancourt, L., M. Du, J. Wang, J. Rollins, and R. Hanau. 2000. Genetic analysis of cross fertility between two self-sterile strains of *Glomerella graminicola*. Mycologia 92:430–435.

White, T.J., T. Bruns, S. Lee, and J.L. Taylor. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. p. 315–322. *In* M.A. Innis, D.H. Gelfand, J.J. Sninsky, and T.J. White (ed.) PCR protocols: A guide to methods and applications. Academic Press, New York.

Wong, F.P., and S.L. Midland. 2007. Sensitivity distributions of California populations of *Colletotrichum cereale* to four sterol demethylation inhibitor fungicides: Propiconazole, myclobutanil, tebuconazole, and triadimefon. Plant Dis. 91:1547–1555.

Wong, F.P., S.L. Midland, and K.A. de la Cerda. 2007. Occurrence and distribution of QoI-Resistant Isolates of *Colletotrichum cereale* from annual bluegrass in California. Plant Dis. 91:1536–1546.

JO ANNE CROUCH
Curriculum Vitae

EDUCATION

- **Ph.D.** **Rutgers University**, New Brunswick, NJ
  Dept. of Plant Biology & Pathology.
  2002- 2008.

- **B.S.** **Rutgers University**, Plant Science, research track,
  with honors, 2002.

PROFESSIONAL APPOINTMENTS

- **Graduate Fellow, Rutgers University:**
  - Geiger Endowment Fellow, 2006 to present.
  - U.S. Environmental Protection Agency STAR Fellow, 2005 to present.
  - The Land Institute Natural Systems Agriculture Fellow, 2004 to present
  - Robert White-Stevens Fellow, 2005-2006.
  - Rutgers University Alberts/Excellence Fellow, 2002-2003.
- **Adjunct Instructor**
  - Kean University, M.S. Program in Biotechnology, *Intro. to Bioinformatics,* Fall 2005
  - Rutgers University, B.S. Program in Biotechnology, *Molecular Genetics Lab*, Spring 2008.
- **Graduate Research Assistant**, Rutgers University, with Dr. Bradley I. Hillman, 2003-2005.
- **Technical Assistant**, Cumberland County College, Dept. of Agriculture & Horticulture, 2000-2002.
- **Landscape Designer,** Scian's Landscaping, Inc., 1992-2000.

AWARDS AND HONORS

- **Mycological Society of America NAMA Memorial Graduate Fellowship**, 2008
- **Rutgers University Graduate Student Research Award**, 2008
- **Theobald Smith Society Graduate Scholarship for Exemplary Achievement in Graduate Study in Microbiology**, Theobald Smith Society of the American Society of Microbiology, 2007
- **Wilbur Runk Graduate Merit Scholarship**, Rutgers Cooperative Extension Service, 2007
- **Geiger Endowment Fellowship**, Rutgers University, Dept. of Plant Biology & Pathology, 2006-2008
- **Mycological Society of America Howard Bigelow Travel Award**: Mycological Association of America, 2006
- **U.S. Environmental Protection Agency (EPA) Science to Achieve Results (STAR) Fellowship for Graduate Environmental Study**: EPA National Center for Environmental Research. Awarded three years of funding in 2005 for proposal entitled: **"**Ecological determinants of population structure and gene flow between sympatric fungal species in the genus *Colletotrichum* from diverse grass communities". 2005-2008
- **Natural Systems Agriculture Graduate Research Fellowship:** The Land Institute. A research grant initially awarded in 2004, renewed in 2005, 2006 and 2007 for research proposal entitled: "Host plant-mediated evolution of the fungal phytopathogen *Colletotrichum cereale* (*graminicola*): Contrasting patterns of adaptation and speciation in diverse grass communities". 2004-2008
- **Executive Women of New Jersey Graduate Merit Award**: Executive Women of New Jersey, 2005
- **Spenser Davis Graduate Research Award**: Dept. of Plant Biology and Pathology, Rutgers University, 2005
- **Robert White-Stevens Graduate Fellowship**: Cook College, Rutgers University, 2005-2006
- **I.E. Melhus Graduate Student Symposium "Honorable Mention",** American Phytopathological Society Epidemiology Committee, 2005.
- **Chaser / National Science Foundation Conference Award**: NSF Regional Grants Conference, Oakland, CA, 2005
- **National Science Foundation K-12 Teaching Fellowship**: Rutgers University Center for Mathematics, 2005, declined
- **Bevier Fellowship "Alternate"**: Graduate School New Brunswick, Rutgers University, 2005.
- **23rd Fungal Genetics Conference Travel Award**, Genetics Society of America, 2005
- **Rutgers Center for Turfgrass Science Scholarship**, Rutgers Center for Turfgrass Science, 2005
- **Steve Johnston Travel Award:** Dept. of Plant Biology and Pathology, Rutgers University, 2005

AWARDS AND HONORS, cont'd

- **Kings Supermarket/Bildner Scholarship**, Cook College, Rutgers University, 2004 and 2007
- **Plant Biology & Pathology Department Travel Award**: Rutgers University, 2004
- **Ralph Geiger Scholarship**: Rutgers Center for Turfgrass Science, 2003 and 2007
- **Peter Selmer Loft Memorial Scholarship:** Rutgers Center for Turfgrass Science; 2002, 2003, 2004, 2005, 2006 and 2007
- **Rutgers Excellence Fellowship**: Rutgers University. 2002-2003
- **Carlos Neyra Travel Award**: Rutgers University, Dept. Plant Biology & Pathology, 2003
- **Plant Science Excellence Award**: Cook College, Department of Plant Science, 2002.
- **FIRST in Floriculture Fran Johnson Scholarship**: Floriculture Industry Research & Scholarship Trust, 2001

PUBLICATIONS

- **Crouch, J.A., Glasheen, B.M., Giunta, M.A., Clarke, B.B. and Hillman, B.I**. (2008) The evolution of transposon repeat-induced point mutation in the genome of *Colletotrichum cereale*: Reconciling sex, recombination and homoplasy in an "asexual" pathogen. Fungal Genetics and Biology 45:190–206.
- **Crouch, J.A., Glasheen, B.M., Uddin, W., Clarke, B.B., and Hillman, B.I.** (2008) Patterns of diversity in populations of the turfgrass pathogen *Colletotrichum cereale* as revealed by transposon fingerprint profiles. In press, Crop Science.
- **Crouch, J.A., Clarke, B.B., and Hillman, B.I.** (2006) Unraveling evolutionary relationships among the divergent lineages of *Colletotrichum* causing anthracnose disease in turfgrass and maize. Phytopathology: 96(1):46-60 (+ 5 pages of online supplemental material; cover article).
- **Crouch, J.A., Clarke, B.B., and Hillman, B.I.** (2005) Phylogenetic relationships and fungicide resistance in *Colletotrichum* isolates from turfgrass in North America. Int. Turfgrass Soc. Res. J. 10: 186-195.
- **Li, H.M., Crouch, J.A., and Belanger, F.C.** (2005) Fungal endophyte *N*-acetylglucosaminidase expression in the infected host grass. Mycological Research 109 (3): 363-373.