THE EVOLUTION OF *LOL*, THE SECONDARY METABOLITE GENE CLUSTER FOR INSECTICIDAL LOLINE ALKALOIDS IN FUNGAL ENDOPHYTES OF GRASSES

A Dissertation

by

BRANDI LYNN KUTIL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2006

Major Subject: Genetics

THE EVOLUTION OF *LOL*, THE SECONDARY METABOLITE GENE CLUSTER FOR INSECTICIDAL LOLINE ALKALOIDS IN FUNGAL ENDOPHYTES OF GRASSES

A Dissertation

by

BRANDI LYNN KUTIL

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Heather H. Wilkinson
Committee Members,	Christine G. Elsik
	Daniel J. Ebbole
	Charles M. Kenerley
Chair of Genetics Faculty,	James R. Wild

December 2006

Major Subject: Genetics

ABSTRACT

The Evolution of *LOL*, the Secondary Metabolite Gene Cluster for Insecticidal Loline Alkaloids in Fungal Endophytes of Grasses. (December 2006)

Brandi Lynn Kutil, B.S., Texas A&M University

Chair of Advisory Committee: Dr. Heather H. Wilkinson

LOL is a novel secondary metabolite gene cluster associated with the production of loline alkaloids (saturated 1-aminopyrrolizidine alkaloids with an oxygen bridge) exclusively in closely related grass-endophyte species in the genera *Epichloë* and *Neotyphodium*. In this study I characterize the *LOL* cluster in *E. festucae*, including the presentation of sequence corresponding to 10 individual *lol* genes as well as defining the boundaries of the cluster and evaluation of the genomic DNA region flanking *LOL* in *E. festucae*. In addition to characterizing the *LOL* cluster in *E. festucae*, I present *LOL* sequence from two additional species, *Neotyphodium coenophialum* and *Neotyphodium* sp. PauTG-1. Together with two recently published *LOL* clusters from *N. uncinatum*, these data allow for a powerful phylogenetic comparison of five clusters from four closely related species. There is a high degree of microsynteny (conserved gene order and orientation) among the five *LOL* clusters, allowing us to predict potential transcriptional co-regulatory binding motifs in *lol* promoter regions. The relatedness of *LOL* clusters is especially interesting in light of the history of interspecific hybridizations that generated the asexual, *Neotyphodium* lineages. In fact, three of the clusters appear to have been introduced to different *Neotyphodium* species by the same

ancestral *Epichloë* species, for which present day isolates are no longer able to produce lolines. To address the evolutionary origins of the cluster we have investigated the phylogenetic relationships of particular *lol* ORFs to their paralogous primary metabolism genes (and gene families) from endophytes, other fungi and even other kingdoms. I present extensive evidence that at least two individual *lol* genes have evolved from primary metabolism genes within the fungal ancestors of endophytes, rather than being introduced via horizontal gene transfer. I also present complementation studies in *Neurospora crassa* exploring the functional divergence of one *lol* gene from its primary metabolism paralog. While it is clear that these insecticidal compounds should convey a selective advantage to the fungus and its host, thus explaining preservation of the trait, this analysis provides an exploration into the evolutionary origin and maintenance of the genes that comprise the *LOL* and the cluster itself.

DEDICATION

This dissertation is dedicated to my family and friends for their endless encouragement and support throughout my education

ACKNOWLEDGMENTS

There is a single light of science, and to brighten it anywhere is to brighten it everywhere. -Isaac Asimov

My heartfelt thanks go out to everyone who has helped me throughout this journey. To Dr. Heather Wilkinson, a true advisor in every sense, thank you for giving me a chance when I had lost faith in myself. I thank Charles Greenwald, Kati Ireland-Stoddard, Heather Thompson, and all of the other promising researchers from the Wilkinson Lab and beyond who have contributed their ideas, hard work and camaraderie. I am grateful to all of the programs and agencies who have provided funding to support my work at various times, including the Department of Plant Pathology and Microbiology, the Interdisciplinary Genetics Program, and the Program for the Biology of Filamentous Fungi (PBOFF), in addition to grants from the NSF and USDA. To my parents and grandparents, thank you for believing in me and supporting me in so many ways. And, most of all, I thank my husband for your understanding and encouragement each and every day of this incredible journey.

NOMENCLATURE

Despite great variability in nomenclature when studying the genetics of different organisms, we have attempted to adhere to a uniform pattern throughout this dissertation as follows:

LOL refers to the genomic DNA sequence encoding the gene cluster;

Lol + or Lol - refers to the phenotype (i.e. the ability to produce loline alkaloids);

lol refers to individual genes comprising the *LOL* gene cluster;

lolD refers to a specific gene within the *LOL* gene cluster;

LolD refers to the amino acid sequence predicted from *lolD*;

LOLD refers to the enzyme produced from the *lolD* sequence.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The Evolution of Secondary Metabolite Gene Clusters

The evolution of secondary metabolite gene clusters is currently receiving interest from scientists studying a broad range of organisms including bacteria (Challis and Hopwood, 2003), plants (Qi et al., 2004; Ober, 2005) and fungi (Walton, 2000; Zhang et al., 2004; Tudzynski, 2005). When considering the evolution of a gene cluster, one must evaluate both the initial formation (***origin***) and the subsequent preservation (***maintenance***) of the cluster. Many mechanisms have been proposed to explain the evolution of gene clusters, however, few of these in isolation can account for both the origin and maintenance of a given secondary metabolite gene cluster. Several of the most long-lived evolutionary models are the Natal model, the Fisher model, the co-regulation model and the selfish operon/cluster hypothesis. The Natal model, based on ideas elegantly proposed by N. H. Horowitz simply states that clusters of paralogous genes may result from tandem duplications (Horowitz, 1945; Lawrence and Roth, 1996). In these clusters, one gene (or occasionally a few genes) is duplicated multiple times, often due to unequal crossover events. These duplicated genes may then diverge to form paralogous versions of the gene. The resulting clustering is merely a position effect.

This dissertation follows the style of Fungal Genetics and Biology.

This process is certainly responsible for the origin of clusters similar to those for imunoglobin genes and rRNA. Indeed, many secondary metabolism gene clusters, particularly in plants, are composed of paralogous genes which have undoubtedly undergone multiple rounds of duplication and diversification (Ober, 2005). However, since the genes formed by this process would all be paralogous, this process cannot explain the formation of clusters composed of nonhomologous genes (such as the biosynthetic gene cluster analyzed in this study).

The Fisher model is more easily applied to fungal secondary metabolite gene clusters (composed of non-paralogous genes). This model begins with a set of co-adapted alleles of different genes (aka. a coadapted gene complex) which provide a greater selective advantage when inherited together (Fisher, 1930). In this situation, recombination between lineages with different alleles would be detrimental for any of the offspring that ended up with a non-optimal combination of alleles. Different lineages harbor different alleles for genes involved in a trait. Particular allele combinations convey substantial fitness advantages. Therefore, recombination among relatively fit lineages with different allele combinations (co-adapted gene complexes) would result in some combinations of alleles that were less fit in the next generation. According to Fisher, under these circumstances selection would favor the most fit combinations of alleles; thus over time one would find linkage disequilibrium among the co-adapted alleles. An extension of this idea is that under strong enough selection and over time actual linkage of the genes (clustering) would be favored in a sexually reproducing

(genetically recombining) population as long as the population maintains genetic variability for the co-adapted gene complex (Stahl and Murray, 1966).

The co-regulation model is also reasonable, postulating that clustering of genes could allow for co-regulation of expression via shared elements for induction or repression (Lawrence and Roth, 1996). Expanded from the operon model in prokaryotes (Jacob and Monod, 1961) the theory does not explain why there would be a multitude of genes which are co-regulated without any physical linkage. Clustering is certainly not a requirement for co-regulation of genes, and furthermore, this model proposes no reasonable mechanism by which clustering may arise. Still, co-regulation might act to maintain a cluster once it is formed, particularly in the case of shared regulatory elements or proteins. Many of the best studied fungal secondary metabolite clusters harbor a pathway specific regulator (e.g. a transcription factor) and most, if not all, the biosynthetic genes in the cluster possess the specific binding sites for this regulator in their promoter regions (Yu and Keller, 2005). Furthermore, recent discovery of global regulators such as LaeA, a regulator of multiple secondary metabolite gene clusters in Aspergillus spp. (Bok et al., 2005; Bok et al., 2006;Yu and Keller, 2005) indicate that chromatin structure may play a significant role in the function of secondary metabolite gene clusters.

After reviewing each of the prevailing theories, Lawrence and Roth proposed the selfish operon model for the evolution of gene clusters in bacteria (Lawrence and Roth, 1996). Incorporating the salient features of each of the three previously described models, in the selfish operon model horizontal gene transfer (HGT) drives the clustering

of functionally related genes, allowing the spread of the co-adapted gene complex. This model is 'selfish' because the clustering of the genes provides a selective advantage not to the individual, but rather to the genes themselves. Segments of DNA that are transmitted horizontally will only persist in new hosts so long as all the components necessary for the trait are present, thus allowing the cluster to convey a selective advantage.

Walton further adapted the selfish operon model to apply to fungi, specifically secondary metabolite gene clusters (Walton, 2000). This is an especially attractive theory for filamentous fungi, which may have the ability to coinfect a host (e.g. plant endophytes) or even fuse together with hyphae of different strains. Clearly HGT would provide an advantage to co-adapted genes that are clustered over those dispersed throughout the genome; however, if we assume that prior to clustering, the co-adapted alleles are dispersed throughout the genome, the probability of each discrete gene being transferred individually via HGT to the same recipient is highly unlikely. Therefore, while HGT may be a mechanism responsible for the dissemination of extant gene clusters between organisms, I find it inadequite to explain the origin of novel clusters.

Other mechanisms implicated in the evolution of individual duplicated genes that make up a cluster or entire duplicated clusters are concerted evolution and the birth and death model of evolution. In the birth and death model of evolution, new genes are initially created by duplication (or even repeated duplication) of an extant gene (Nei et al., 1997). Gene copies are likely to experience selection with reduced functional constraints and may diverge to different allelic forms or even completely different

functions resulting in subfunctionalization or neofunctionalization. Subfunctionalization essentially divides the responsibilities of multiple gene functions among the independent paralogous forms; whereas neofunctionalization maintains one gene copy with all the original gene functions and allows the duplicate(s) to diverge to take on an entirely new function. Some of these paralogous forms may be maintained by the organism while others may accumulate deleterious mutation and become nonfunctional or be deleted (Nei et al., 1997). Concerted evolution uses gene conversion or unequal crossing-over between misaligned paralogous genes to explain how these genes present multiple times in a single genome may be more conserved than expected or share common mutations which have arisen after the original duplication of a cluster (Nagylaki, 1984; Ohta, 1990). While less likely to play a role in the evolution of an entire secondary metabolism gene clusters (composed of nonhomologous genes), both of these theories may be implicated in the evolution of gene clusters composed of tandem duplications, gene clusters represented multiple times in a single genome and the individual genes found in (nonhomologous) biosynthetic gene clusters. Generally speaking, the dispensible nature of secondary metabolite gene clusters makes them of particular interest for understanding the evolution of novel traits and functions. That is, because the traits these clusters encode are not essential (ie. loss of the trait is not lethal to the organism) secondary metabolite gene clusters are not subject to the same stabilizing selection as genes for traits that are absolutely required for survival or reproduction.

In this thesis I explore the evolution of a fungal secondary metabolite gene cluster, *LOL*, which is associated with the production of loline alkaloids in fungal grass-

endophytes (*Epichloë* and *Neotyphodium* species). Specifically, I sought to determine to what degree each of the evolutionary mechanisms proported to be involved in evolution of secondary metabolite clusters (co-regulation, HGT, gene duplication, functional divergence, synteny) was evident in the analysis of *LOL*.

Host Fitness Enhancements Associated with Grass-endophyte Mutualisms

Epichloë festucae is a mutualistic fungal endophyte that provides fitness enhancements to its grass host. Grasses associated with endophytes of the related genera *Epichloë* (sexual) and *Neotyphodium* (strictly asexual) show increased competitiveness (Clay, 1989; Marks et al., 1991) and tolerance to biotic and abiotic environmental stresses including: insect herbivory (Clay, 1989; Dahlman et al., 1997; Latch, 1993; Popay and Rowan, 1994; Siegel et al., 1990), mammalian herbivores (Lacey, 1991; Lyons et al., 1986; Raisbeck et al., 1991), nematodes (Kimmons et al., 1990), fungal diseases (Funk and White, 1997), drought (West, 1994) and various soil stresses (Belesky and Malinowski, 2000). These enhancements are associated with the production of four different classes of alkaloids: peramine is a feeding deterrent to insects (Rowan and Gaynor, 1986); ergot alkaloids and indolediterpenes are active against insects and vertebrates (Clay and Cheplick, 1989; Lacey, 1991; Raisbeck et al., 1991; Rowan and Latch, 1994); loline alkaloids (Fig. 1) are insecticidal to a broad range of insects (Bush et al., 1997; Dahlman et al., 1997; Johnson et al., 1985; Wilkinson et al., 2000;Yates et al., 1989) with no known effects on mammals.

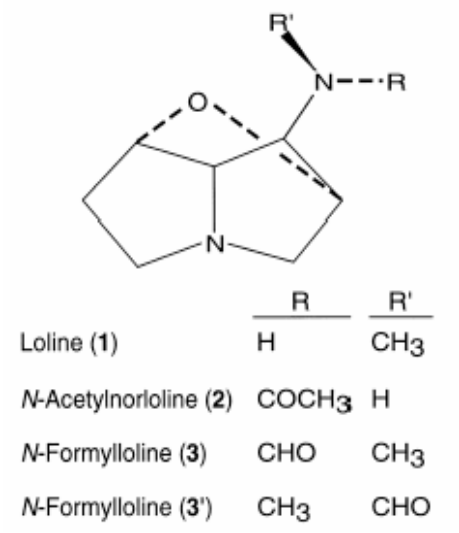

Fig. 1. Structures of some loline alkaloids.
(Figure adapted from Blankenship et al., 2001)

Loline alkaloids are also circumstantially implicated in the enhancement of drought tolerance for the following reasons: 1) drought and heat increase loline accumulation in tissues responsible for regrowth (Bush et al., 1997); 2) lolines are water soluble and accumulate to such high levels (0.1-2% of the dry weight of the plant) to expect effects on the plant's osmotic adjustment; 3) drought tolerance has been demonstrated only by symbiota that produce lolines and not by symbiota that produce other alkaloids but not lolines (Arechavaleta et al., 1989; Belesky et al., 1989; Malinowski, 1995; Malinowski et al., 1997; Read and Camp, 1986).

Genetic Analysis of Loline Alkaloid Production in *Epichloë festucae*

Previously, Wilkinson et al. identified two natural isolates of *Epichloë festucae* that varied qualitatively in their ability to produce loline alkaloids (Lol+ vs Lol-) (Wilkinson et al., 2000). These two natural isolates were mated, and analysis of lolines expression in their progeny indicated that the lolines polymorphism segregated as a single locus, *LOL*. Furthermore, they identified an AFLP which segregrated with the ability to produce lolines (Wilkinson et al., 2000). In order to map this AFLP and clone the cluster a BAC library was constructed from a lolines-producing progeny isolated from the original *E. festucae* cross. No genome sequence was available for *E. festucae* or other closely related endophytes. While *LOL* was initially identified by map-based cloning in *E. festucae*, it should be noted that discovery of many of the individual *lol* genes occurred concomitantly based on efforts in this study of *E. festucae* and work by Dr. Christopher Schardl's laboratory at the University of Kentucky analyzing the closely related species *N. uncinatum*. Sequence data was freely shared between the two research teams, allowing for more rapid progress in both systems.

Overview

In the following studies, I present a multifaceted exploration of the evolution of the fungal secondary metabolite gene cluster *LOL* responsible for production of loline alkaloids. An AFLP that segregated with the ability to produce lolines was mapped to several clones in our *E. festucae* BAC library. Initial characterizations of these BACs via subcloning and sequencing identified several genes with homology to the genes

predicted to be involved in the biosynthesis of loline alkaloids. Therefore, one of these BACs (BAC 12-O9) was subjected to shotgun sequencing and assembled to reveal two very different regions. The T7 end of the BAC contains a 40 kb region which flanks *LOL* but does not appear to be involved in LOL biosynthesis. It is present in both Lol+ and Lol- parental isolates and contains many housekeeping genes which are conserved across filamentous fungi and, in some cases other kingdoms as well. Since so little *E. festucae* sequence was previously available, sequence from this region provides an ideal opportunity to explore the levels of microsynteny between *E. festucae* and model fungal systems. The moderate levels of conservation of microsynteny observed between the *E. festucae* and other filamentous fungi are encouraging as we further explore evolution and conservation within individual gene families in the next chapters. In addition to evaluating the microsynteny of the genes flanking one side of the *LOL* gene cluster in *E. festucae* relative to orthologs in model filamentous fungi, Chapter II describes the construction and characterization of our *E. festucae* BAC library.

At the sp6 end of BAC 12-O9 a 50kb region unique to the lolines-producing parental isolate contains eight orfs predicted to be part of the *LOL* gene cluster. This region is characterized in Chapter III. Since the region flanking *LOL* shows reasonable conservation of gene order and orientation across the Ascomycota, we hypothesized that the conservation of microsynteny of the *LOL* gene cluster would be even more evident in the closely related lolines-producing endophyte species examined. Chapter III describes both i) the map-based-cloning and characterization of the *LOL* gene cluster in *E. festucae* and ii) the comparison of sequences of five *LOL* clusters from four closely

related endophyte species (*E. festucae, Neotyphodium uncinatum, N. ceonophialum* and *Neotyphodium* sp. PauTG-1). This study explores the recent evolution of *LOL*, after its formation in or introduction into the ancestors of modern day endophytes. The endophytes that produce lolines are found in the highly diversified family Clavicepitacae. However, it should be noted that some isolates of the two lolines-producing genera identified thus far, *Epichloë* (sexual) and *Neotyphodium* (asexual), have been identified as over 99% identical at the DNA level. The asexual *Neotyphodium* species used in this study are generally accepted to be the product of interspecific hybridizations of *Epichloë* species. *Epichloë festucae* is the only sexual species identified thus far which is able to produce loline alkaloids, and even in natural isolates of this species lolines expression is rare. We hypothesize that the *LOL* clusters found in the *Neotyphodium* species can each be traced back to one of the ancestral *Epichloë* species that contributed to the genome of that *Neotyphodium* species. *E. festucae* is implicated as an ancestor to *N. coenophialum* and was initially hypothesized to be the contributor of *LOL* in that species. However, since *E. festucae* is not implicated as an ancestor for *N. uncinatum* (which has two copies of *LOL*) or *Neotyphodium* sp. PauTG-1, we seek to discover if the *LOL* cluster(s) in those two species are highly similar (indicating origin from a shared ancestor, *E. typhina*) or relatively divergent (indicating *LOL* may have been derived from different ancestors). The results described in the study answer each of these hypotheses rather definitively and, in some cases, surprisingly.

Chapter IV addresses the events leading to the origin of *LOL*. Specifically, the chapter explores the evolutionary origin of two individual *lol* genes that appear to be

unique to the *LOL* cluster. While a region homologous to the *LOL* cluster is not present in any sequenced genome, BLAST comparison and intron analysis of individual *lol* genes indicates that these genes are most closely related to fungal gene families. Two *lol* genes, *lolD* and *lolC*, are paralogous to genes involved in amino acid biosynthesis. A multi-level gene family analysis was undertaken to determine the most likely origins of each of these genes. These studies present evidence for duplication of primary metabolism genes within the genome of a common ancestor to the loline producing endophytes and subsequently neofunctionalization resulting in the evolution of novel *lol* genes within two fairly well understood gene superfamilies. *lolD* appears to have evolved from the fungal ornithine decarboxylase gene family. The amino acid translation predicted from *lolC* shares homology with O-acetylhomoserine sulfhydrylase (OAHsh), a member of the cysteine methionine metabolism PLP-dependent (Cys/Met Metab PP) gene family. All of these findings support our hypothesis that the biosynthesis of loline alkaloids is a trait that has originated within the common ancestor to modern-day *Epichloë* and *Neotyphodium* endophytes.

Chapter V describes the initial functional analysis of *lolD*, one of the genes featured in the evolutionary analysis (Chapter IV). In that study, I discovered that *lolD* likely evolved from *ornithine decarboxylase* (*odc*). Thus in Chapter V, I set out to determine whether or not the secondary metabolism *lolD* produces a protein which can functionally complement ODC, the product of its primary metabolism paralog. Exploration of function between homologs is certainly important in evolutionary analysis because homologs that maintain overlapping functions are under different

evolutionary pressures than those that have assumed different functions (neofunctionalization). While LOLD is likely to perform a function similar to the primary metabolism ODC (decarboxylation of a substrate similar to ornithine, see Chapter V Introduction), we predicted that it should be sufficiently diverged that LOLD cannot complement a nonfunctional ODC. To evaluate this hypothesis, I complemented a *Neurospora crassa* strain carrying a non-functional *odc* allele with the *lolD* gene from *N. coenophialum* and, as a positive control, an *odc* gene from *Epichloë festucae*. The results of these experiments support my hypothesis that *lolD* has diverged enough from *odc* to not restore the ODC function in *N. crassa*. However, the *E. festucae odc* gene did restore the wildtype phenotype for the *Neurospora* mutant. In addition to being a successful positive control, these results indicate that the model organism *N. crassa* may serve well in future investigations of endophyte genes.

Finally Chapter VI summarizes the findings of this dissertation project. Specifically, the chapter places the work in context of the current state of understanding of evolution secondary metabolite clusters, as well as setting forth suggested future directions for research involving the evolution of the loline alkaloid gene cluster.

CHAPTER II

CONTIG ASSEMBLY AND MICROSYNTENY ANALYSIS USING A BACTERIAL ARTIFICIAL CHROMOSOME LIBRARY FOR *Epichloë festucae,* A MUTUALISTIC FUNGAL ENDOPHYTE OF GRASSES*

Introduction

One frequently asserted promise of comparative genomic approaches among fungi is that it will be possible to elucidate the genetic mechanisms associated with different ecological lifestyles (e.g. saprophyte, pathogen, mutualist) (Tunlid and Talbot, 2002). Presumably, since each of these different lifestyles is represented across a wide array of taxonomic groups they have evolved independently many times (Berbee, 2001; Tunlid and Talbot, 2002). Thus, fungal species with different ecological lifestyles are likely using a similar cadre of genes to execute each of these different interaction outcomes. It is reasonable to expect that discerning the basis for ecological differentiation requires examining variation associated with closely related species exhibiting different lifestyles. Attempts to capitalize on the wealth of forthcoming

*Reprinted from Fungal Genetics and Biology, Vol **41**, Brandi L. Kutil, Gang Liu, Julia Vrebalov, and Heather H. Wilkinson, "Contig assembly and microsynteny analysis using a bacterial artificial chromosome library for *Epichloë festucae,* a mutualistic fungal endophyte of grasses", Pages **23-32**, Copyright (2004), with permission from Elsevier.

genomic data from filamentous fungal systems will require focused efforts to include ecologically diverse systems that have diverged recently relative to the available models.

Epichloë festucae, a fungal endophyte of cool season grasses in the genera *Festuca* and *Lolium* (Leuchtmann et al., 1994), is being developed as a model for a fungal mutualist of plants (Schardl, 2001; Wilkinson et al., 2000). This heterothallic ascomycete is closely related to most of the pathogenic and saprophytic fungal species currently available or slated for sequencing within the near future (http://www-genome.wi.mit.edu/annotation/fungi/fgi/). Furthermore, with its small genome (29Mb; Kuldau et al., 1999) and the ability to perform crosses under laboratory conditions (Wilkinson et al., 2000) *E. festucae* is arguably more genetically tractable than other well established mutualistic fungi (e.g. AM endomycorrhizae, Tunlid and Talbot, 2002).

Mutualistic *Epichloë* spp. and their anamorphic relatives (*Neotyphodium* spp.) enhance the fitness of their hosts by increasing resistance to a wide variety of biotic and abiotic stresses (Bush et al., 1997; Schardl, 2001). Well established host fitness enhancements include: increased resistance to herbivory by mammals (Lacey, 1991; Lyons et al., 1986; Raisbeck et al., 1991) and insects (Clay, 1989; Dahlman et al., 1997; Latch, 1993; Popay and Rowan, 1994; Siegel et al., 1990; Wilkinson et al., 2000), increased resistance to drought stress (West, 1994), increased competitive ability (Clay, 1989; Marks et al., 1991), increased resistance to pathogens such as nematodes (Kimmons et al., 1990) and fungal diseases (Funk and White, 1997) and greater tolerance to stressful soil conditions (Belesky and Malinowski, 2000). Since many endophytes are mutualists of agriculturally important forage grasses (Schardl, 2001),

development of the *E. festucae* model should contribute to ongoing efforts to improve endophyte traits that affect forage quality (Panaccione et al., 2001). Furthermore, based on the extensive phylogenetic analysis of mutualistic and more pathogenic endophyte species (Craven et al., 2001a; Craven et al., 2001b; Schardl et al., 1994; Tsai et al., 1994) this system may represent the best opportunity to discern the evolutionary history associated with adaptive shifts along the symbiotic continuum between pathogen and mutualist (Wilkinson and Schardl, 1997; Clay, 1988). Furthermore, *E. festucae* is a member of the family *Clavicipitaceae*, which has an especially well established phylogeny (Schardl et al., 1994; Spatafora and Blackwell, 1993; Sung et al., 2001) exclusively comprised of obligately biotrophic species, forming associations with a diversity of hosts (plants, insects and fungi) from three eukaryotic kingdoms. Ultimately, development of a model within this ecologically diverse group will provide opportunities to address the evolution of novel host affiliations.

To facilitate genome analysis we have constructed a BAC library (Kutil et al., 2004). To assess the utility of this resource for map-based cloning and ultimately whole genome mapping and gene discovery, we conducted experiments to determine both genome coverage and amount of repetitive DNA present in the library. Both whole genomic DNA and 6 single copy DNA segments were used as probes against the library. To assess the utility of the library for contig assembly we physically mapped several BACs identified in screens with one of the single copy genes and extended the initial contig by chromosome walking a single step in both directions. Furthermore, since any degree of gene discovery and functional prediction in this system is expected to be

facilitated by the availability of genome sequence from model ascomycetes, we used sequence available from the end of one BAC to directly test the hypothesis that there would be conservation in microsynteny between *E. festucae* and the more closely related publicly available fungal genome sequences. Specifically, we compared 32 kb of sequence mapped to a 43kb region in *E. festucae* to sequence from *Neurospora crassa*, *Magnaporthe grisea*, *Aspergillus nidulans*, and *Cryptococcus neoformans*.

Materials and Methods

Fungal culture conditions for high molecular weight DNA extraction

To generate cultures for protoplasting and high molecular weight (HMW) DNA extraction, one standard sized potato dextrose agar plate was streaked with conidia from a pure culture of *E. festucae* isolate 1035.30 (Wilkinson et al., 2000) and allowed to fill with mycelia over 8 days at 25ºC. The entire contents of this plate were then placed in a 250 mL blender (Waring) with 30 mL of sterile distilled water and blended on high for two 10 second pulses. A 25 mL volume of the resulting slurry was used to inoculate 100mL of potato dextrose broth (PDB). This culture was grown shaking at 100 rpm for 3 days at 25ºC and its entire contents were blended in a similar manner. 25 mL of this homogenized material was used to inoculate each of the two final 1000mL PDB cultures. The final cultures were grown shaking for 3 days at 25ºC. The fungal material was collected by centrifugation at 10,000 rpm for 5 minutes in 250 mL centrifuge tubes. The spent medium was poured off and the fungal biomass was rinsed with sterile water,

recentrifuged, and the pellets were collected and stored on ice until protoplasting. Protoplast preparation for HMW genomic DNA was performed as descried in (Diaz-Perez et al., 1996).

Partial digestion and size selection of HMW DNA

HMW Genomic DNA was partially digested with *Hin*dIII then separated by pulsed field gel electrophoresis (PFGE) for three rounds of size selection. In the first two size selections, the pulsed field gel was run with a 90 s pulse at 160 V and 11°C for 18 h and DNA ranging from 200 kb to 400 kb was excised from the gel for use in the next size selection. The final size selection gel was run using a 6 s pulse at 150 V and 11°C for 11 h. The compressed band representing DNA fragments greater than 150 kb was excised and the DNA was released from the agarose by electroelution. The agarose gel slice containing the > 150 kb DNA was fragmented with a razor blade and the resulting pieces placed in dialysis bags (Gibco BRL). Electroelution was carried out for 2 h at 200 V with a 90 s pulse at 11°C. Eluted DNA was quantified on an agarose gel.

BAC library construction and storage

The size-fractionated 1035.30 DNA was ligated with the vector in an approximately 3:1 molar ratio. *Hin*dIII digested and dephosphorolated pBeloBAC 11 (Shizuya et al., 1992) was used for library construction. Transformations were performed by electroporation using Gene Hogs electrocompetent cells (Research Genetics). Recombinant BAC-containing colonies were grown in 50 μl of 2x TY broth

in 384-well microtiter plates. Plates were incubated for 24 h at 37°C, then the library was replicated into sets of 384-well microtiter plates. Following incubation, freezer medium (360 mM K_2HPO_4, 132mM KH_2PO_4, 17 mM sodium citrate, 4mM $MgSO_4$, 68mM $(NH_4)_2SO_4$, 44% glycerol) was added and the plates were sealed and stored at -80°C. For screening, the library plates were replicated onto 4 high-density nylon membranes with a Biomek 2000 robot (Beckman-Coulter). Each 8 x 12 cm^2 membrane contains clones from four 384-well plates spotted in duplicate.

Preparation and analysis of BAC clones

To detect the average insert size, BAC DNA extraction was performed in a 96-well format using a modified alkaline lysis method (Klein et al., 1998). *Hin*dIII- and *Not*I-digested BAC DNAs were separated by agarose or PFGE gel electrophoresis, respectively. The DNA fragments digested with *Not*I were separated in TAE buffer at 10°C and 6 V/cm with 90 s pulse for 20 h. The conditions for agarose gel electrophoresis were 40V 22 h runs with recirculating TAE at 10°C on a 20 x 25 cm^2 Gator A2 electrophoresis system from Owl. Gels were 170 ml of 1% agarose (Seakem) and were stained in 170 ml of a 1:10,000 dilution of SYBR Gold in TAE. Images captured on an Alpha Imager using v5.5 were exported as TIFF files and processed through Image v3.10 (http://www.sanger.ac.uk/Software/) and FPC V6 (Soderlund et al., 2000; http://www.genome.clemson.edu/fpc/).

For Southern-blot analysis, agarose gels were blotted onto Hybond N+ membranes (Amersham, USA). Hybridization was carried out either a) in phosphate

hybridization buffer (0.125 M $NaHPO_4$; 7% SDS; 1 mM EDTA) at 60°C for 16 h, and washed initially in 1x SSC-1% SDS for 15 min then 0.1x SSC-0.1% SDS for 15 min at room temperature, or b) in 50% deionized formamide in the phosphate buffer and hybridized at 45°C, and the first wash was repeated twice at 45°C, and the final wash was done at 65°C.

To assess the proportion of repeat DNA in the library, 1035.30 genomic DNA was used as a probe (Zhu et al., 1997). Briefly, 50 ng genomic DNA was digested with *Hin*d III at 37°C for 3 h then divided into 2, 25ng reaction mixtures and labeled using the Multiprime DNA labeling kit (Amersham Pharmacia biotech) in the presence of ^{32}P-dCTP.

To test the representation of single copy sequences in the library, 6 single copy genes were used to probe the library membranes (Table 1). Probes for *act1*, *tef1*, and *tub2* were generated directly from PCR products (Table 1). The *odc* probe was a PCR product cloned into the pCR4-TOPO vector, while the clones of *ms* and the aflpB marker were in pBluescript.

odc Contig assembly

Contig assembly surrounding a putative ornithine decarboxylase gene (*odc*) ortholog first involved development of degenerate primers to amplify the gene. Torres-Guzman et al., 1996 designed degenerate primers (SP1 with ASP1 or ASP2) to amplify *odc* from fungi. Those primers failed to amplify a product from *E. festucae* DNA. We

Table 1
Representation of single copy sequences in the *E. festucae* BAC library

Locus	Description	Number of BAC clones	Source
act1	gamma-actin	48	Fidel et al. 1988[a]
aflpB	AFLP marker	3	Wilkinson et al. 2000[b]
ms	methionine synthase	50	unpublished data[b]
odc	ornithine decarboxylase	8	this study[b]
tef1	elongation factor 1-alpha	33	O'Donnell et al. 1998[a]
tub2	beta tubulin	25	Schardl et al. 1997[a]

[a] Both the sequences of primers used for amplification of portions of the *act1*, *tef1* and *tub2* genes, and the establishment that these genes are single copy in the *E. festucae* genome are based on Craven et al. (2001a).
[b] Southern-blot analysis using these loci as probes against *E. festucae* genomic DNA indicate they are also single copy (data not shown)

designed new primers, ODC Up 5'- GCATCATCTATGCCCAGCCNTGYAARAC and ODC Low 5'-CGATGCCGTCGCAGGTNGGNCCCCA, based on the aligned *odc* sequences from *Tapesia yallundae*, *Neurospora crassa*, *Paracoccidioides brasiliensis*, *Phaeosphaeria nodorum*, and *Coccidioides immitis*. The five amino acid sequences were aligned by ClustalW (Thompson et al., 1994), then blocks were generated by GIBBS (Lawrence et al., 1993), and primers were designed by CODEHOP (Rose et al., 1998) (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html). Primer clamps were altered manually to be more likely to match the *N. crassa* gene as it was the most closely related ascomycete with sequence available for *odc.* The new primers amplified products of expected size from genomic DNA of *Epichloë festucae, Trichoderma virens*, and *Neurospora crassa.* Specifically, conditions involved using AmpliTaq Gold with 3 mM $MgCl_2$ and 1μM each primer with the following cycling conditions: 94°C denaturation for 30 s, 55°C annealing for 30 s, and 72°C extension for 1 min, repeated for 42 cycles,

followed by a final 10 min extension. The approximately 750 nucleotide band from the *E. festucae* amplification was extracted, cloned into the pCR4-TOPO cloning vector (Invitrogen), and sequenced. The BACs identified with this probe were subjected to both fingerprint analysis and BAC end-sequencing. DNA extraction from individual BACs for end sequencing, Southern-blot analysis, or locus-specific restriction mapping followed the Qiagen midi-prep kit protocol for 500 mL cultures (QIAGEN, CA). Ends of BAC clones were sequenced using the dye terminator cycle sequencing method (GeneAmp PCR Systems 9600). Automated sequencing was performed with an Applied Biosystems 373 XL DNA sequencer and data analysis was performed with Sequencher 3.1 software (Gene Codes Co., Ann Arbor, Mich.).

Subcloning and sequencing a BAC

BAC 14-O9 which contains a 100kb genomic fragment spanning a region between the methionine synthase and aflpB markers (Table 1), was subcloned and sequenced (CA&ES Genome Facility, University of California, Davis). High quality BAC DNA was isolated for use in shotgun library construction (Qiagen midi prep). The DNA was sheared using a Hydroshear (GeneMachines) in order to create truly random clones of a particular size (2-4 kb is the optimal size for both cloning and generating the best coverage using bidirectional sequencing). Sheared DNA was cloned into the TOPO Shotgun Cloning Kit (Invitrogen) and inserts were sequenced from both directions.

Sequence data were trimmed and assembled using Sequencher 3.1 software Sequence data used in the microsynteny analysis correspond both to the T7 end of BAC

14-O9, closer to the methionine synthase marker (*ms*), and to end sequence data from other BACs associated with the region. Altogether the 32 kb of sequence assembled into seven contigs spanning approximately 43 kb. The seven contigs were mapped and anchored based on standard approaches including restriction digest of BACs and end-sequencing of clones.

Microsynteny analysis

We assessed the degree of conservation of microsynteny between *E. festucae* and some model filamentous fungi. Initially, sequence data from the seven *E. festucae* contigs (Genbank accession **AY365418**) were analyzed by BLAST searches against the predicted proteins (blastx) and/or genomic sequence data (tblastx) for *Neurospora crassa*. Contigs were also submitted to FGENESH (www.softberry.com/berry.phtml) to predict ORFs using the *Neurospora crassa* as the training set. Subsequently, the 13 *E. festucae* ORFs discerned in these ways were compared to *N. crassa, M. grisea, A. nidulans*, and *C. neoformans* using blastp and tblastn. Further, direct comparisons of neighboring ORFs from these microsyntenic regions for the fungi most closely related to *E. festucae* (*N. crassa* and *M. grisea*) were conducted using blastp and blastx. Size, orientation and other aspects of colinearity for ORFs involved in microsynteny analysis were also confirmed by comparison to the feature map and physical and genetic maps of the *N. crassa* and *M. grisea* available through the Center for Genome Research at the Whitehead Institute (http://www-genome.wi.mit.edu/).

Results and Discussion

A library of 6,144 BAC clones was constructed from the grass endophyte *Epichloë festucae* isolate 1035.30. The library is stored as individual colonies in 384-well microtiter plates and available for screening on four high-density membranes. Evaluations of the distribution of insert sizes as well as the mean insert size are necessary for the characterization of library coverage. Initial size estimates calculated based on *Not*I digests of 25 BAC clones revealed insert sizes ranging from 65 kb to 235 kb with an average size of 125 kb (Fig. 2A). For a more thorough estimation of insert size and distribution, images of *Hin*dIII digests of 176 BACs separated on agarose gels were analyzed using Image to help call bands and sizes (Fig. 2B). Of the 176 random BAC clones analyzed four appeared to have no insert. The 172 remaining BACs ranged in size from 29 to 148 kb with the mean insert size of 87.4 kb (Fig. 2C). We expect the average insert size is actually larger than this estimate because it was not possible to fully resolve the many bands smaller than 1.6 kb on these agarose gels.

Based on an estimated size of 29 Mb per haploid genome for *E. festucae* (Kuldau et al., 1999), and an average insert size of at least 87 kb, this library represents greater than 18.4 genomic equivalents. Based on the equation $F = (1-(I/G))^N$ where N represents the number of clones with an average insert size of I from a genome of size G, the theoretical likelihood of failure, F, to find any given sequence represented in the library is less than 10^{-8} (Clark and Carbon, 1976). In reality, some regions of DNA do not clone as well as others. However, it has been established that a wide range of BAC inserts sizes, as is the case for this library, results in a higher probability of representation than a

library with similar average insert size but a small range of insert sizes (Zhang and Wing, 1997).

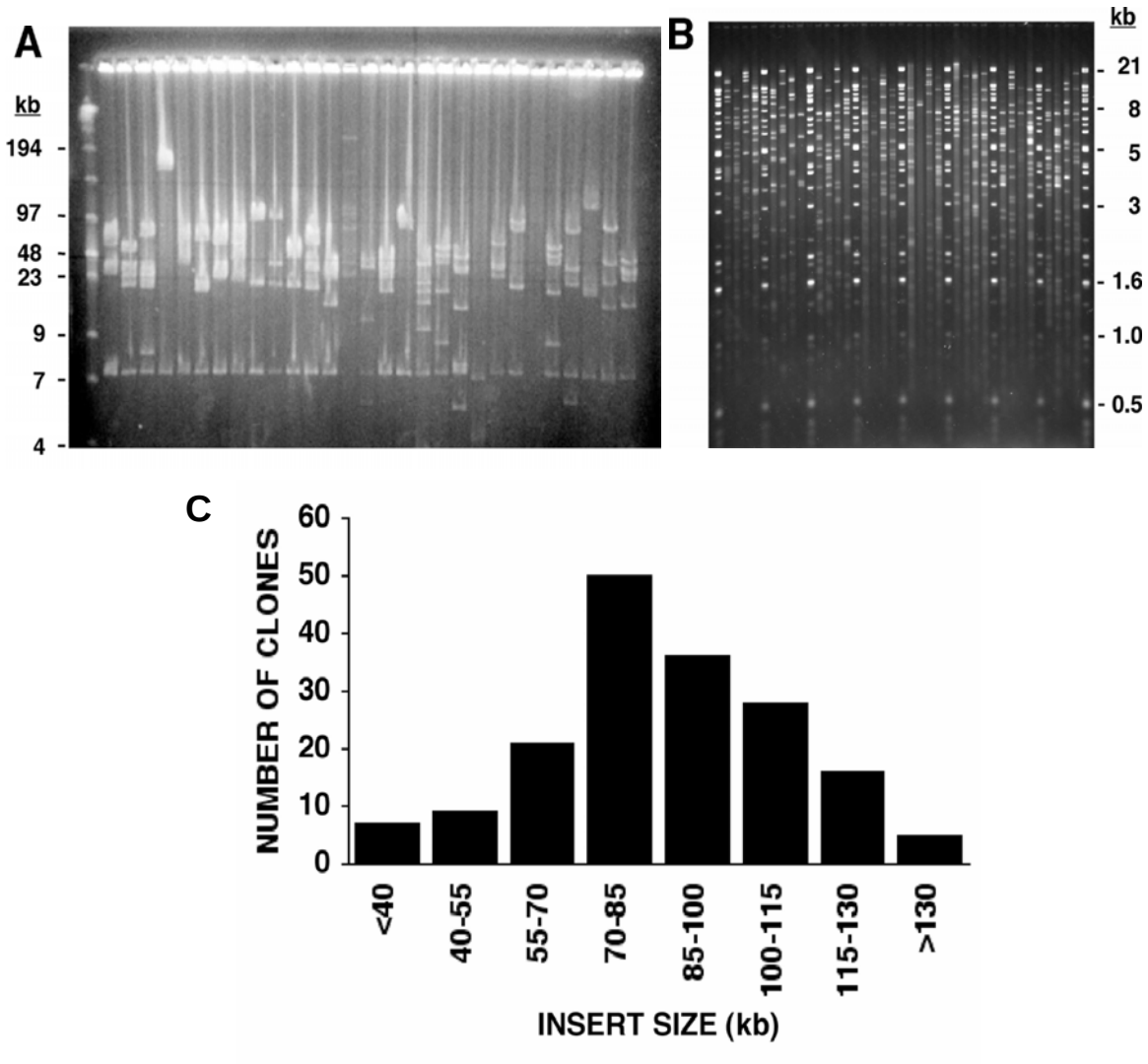

Fig. 2. Insert size determination for randomly selected BAC clones from the *E. festucae* library. (A) *Not*I digestions were separated by pulsed-field gel electrophoresis. The average insert size of this initial sample of BACs was 125 kb. (B) *Hin*dIII digestions of BAC clones were separated by agarose gel electrophoresis. Molecular weight marker is shown in every fifth lane. (C) Insert size distribution of 172 BAC clones as analyzed in panel B. The average insert size for the *Hin*dIII digested BACs analyzed by agarose gel separation was 87 kb.

By probing the BAC colony membranes with total genomic DNA from *E. festucae* 1035.30, we hoped to discern the proportion of clones in the library that contain repetitive DNA. It is reasonable to expect that clones containing single-copy DNA are more likely to show no hybridization signal, while clones containing moderate to highly repetitive sequence should yield a hybridization signal (Zhu et al., 1997). Thus, the hybridization signal of each of the 6,144 BAC clones was scored as either present or absent. Specifically, a clone with a hybridization signal for both the duplicate copies that was above background for the membrane was categorized as containing repetitive DNA. Based on these scores 20% of the BACs in this library contain repetitive DNA (Fig. 3A). To confirm the accuracy of this assay, eighteen clones putatively assigned as containing repetitive DNA and eighteen clones scored as containing no repetitive DNA were randomly selected. DNA from each clone was isolated, digested with *Hin*d III, separated on an agarose gel (Fig. 3B), and analyzed by Southern blot using *E. f*estucae 1035.30 total genomic DNA as a probe (Fig. 3C). (Note, one clone from each of these sets of 18 was eliminated from this analysis because it appeared degraded on the gel; Fig. 3B). All but one of the 17 clones that were assigned to the non-repetitive category showed low or no hybridization signal, while all the clones expected to contain repetitive sequence show high intensity multi-bands (Fig. 3C). Thus, our ability to identify repetitive and non-repetitive clones in the library based on hybridization signal was relatively accurate. Interestingly, among the thirty BACs we have analyzed from those indicated to contain repetitive DNA, we have found a recurrent pattern of seven bands present in five of the

highest signal intensity clones, which may indicate the presence of a common repetitive element (data not shown).

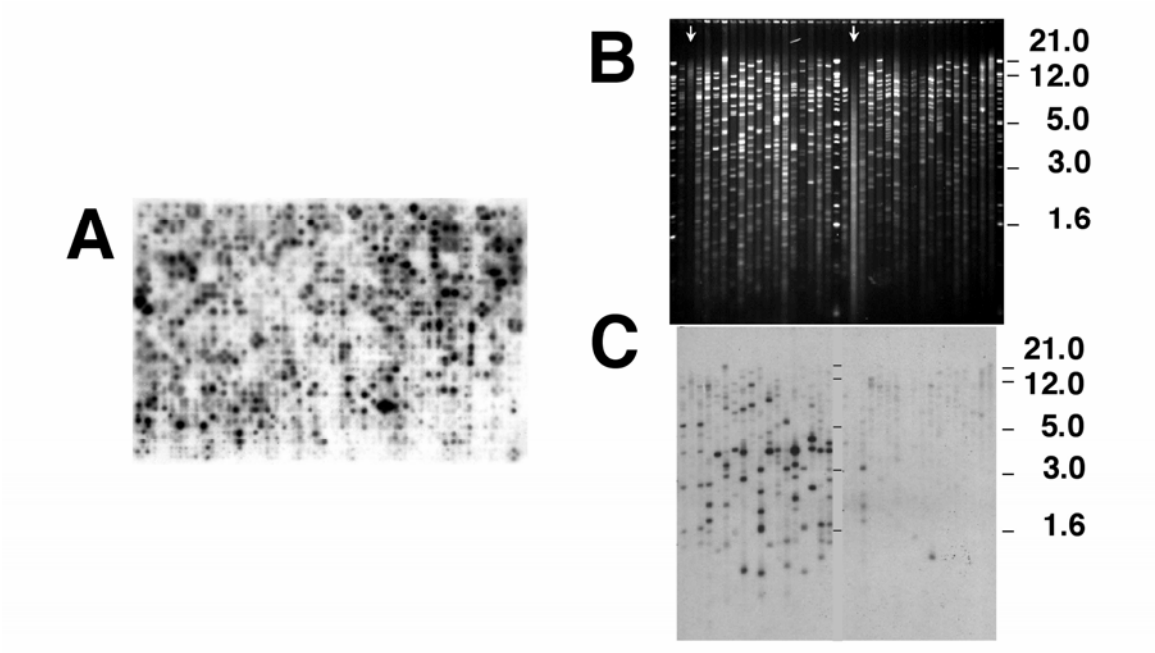

Fig. 3. Determination of repetitive DNA content for the *E. festucae* BAC library. (A) Radioactively labeled genomic DNA was used as a probe against the four high-density membranes of the BAC library. One membrane is shown in this panel. Based on presence or absence of a hybridization signal, repetitive DNA was indicated in 20% of the 6,144 BAC clones. (B) *Hin*dIII digests of BACs randomly selected from the "repetitive DNA containing" (left side) and the "non-repetitive DNA containing" (right side) categories. (C) Southern blot of the gel in panel B was subsequently hybridized with radioactively labeled genomic DNA. The pattern supports the assignment of these clones to particular repetitive DNA categories. The arrows indicate degraded BAC DNA. The remaining 17 repetitive clones have many bands showing high intensity, while most of the non-repetitive clones show little or no signal.

The relatively low percentage of clones containing repetitive DNA (20%) is promising for the utility of the BAC library. Whether or not this low level of repetitive DNA within BAC clones is indicative that repetitive elements are less abundant in *E. festucae* than in some other fungal species remains a question. By way of comparison, using an identical strategy to evaluate the BAC library of the plant pathogen *M. grisea* Zhu et al. (1997) estimated that 74% of the clones contained repetitive DNA. This is not surprising since *M. grisea* is known to harbor an abundance of repetitive elements (Farman et al., 1996; Kachroo et al., 1994; Kachroo et al., 1995; Kang et al., 1995) However, it should be emphasized that the proportion of clones containing highly repetitive sequences may not necessarily reflect the genomic structure. For example, in the more-distantly related oomycete *Phytophthora infestans*, at least 50% of the genome is comprised of repetitive sequences (Judelson and Randall, 1998) but only 30% of the clones in the *Phytophthora* BAC library indicate the presence of repetitive sequences (Randall and Judelson, 1999).

To evaluate the coverage of the library, we screened the high density BAC membranes with three clones and three PCR products from single copy sequences. As indicated in Table 1, all sequences tested were represented in the library. These single-copy regions varied in representation with 3 to 50 library clones harboring each sequence. This supports an estimate of greater than 18X genome coverage. The presence of each of the six single-copy sequences in the library demonstrates that this library is robust.

Exploration of the region around the AFLP marker which was represented only three times in the library revealed several open reading frames with homology to genes encoding reverse transcriptases (See " *Sequences with similarity to pol elements mapping near LOL in E. festucae",* in Chapter III, Results) and some regions of DNA with greater than 70% AT content. Thus, the presence of this AFLP in three BACs is encouraging, indicating that regions classically considered difficult to clone may be represented in this BAC library.

As a test of our ability to use the library in contig assembly, the BACs identified by the *odc* probe (Table 1) were pursued further. FPC analysis of gel images from *Hin*dIII digests of these 8 BACs yielded a 110 kb contig initially. We sequenced the ends of each BAC and then used the unique end sequences from three BACs to generate six additional probes (A through F in Fig. 4), which were hybridized both to these original 8 BACs and to the entire library. Probe C hybridized to precisely the same BACs as the original *odc* probe. Probes D, E, and F hybridized to highly overlapping sets of BACs, containing 38, 34, and 45 clones respectively. Probes A and B hybridized weakly to over 100 BACs each with fewer than half recognized by both probes, even though they map less than 7 kb apart. We subjected 32 of the BACs identified by the six probes to fingerprinting. Not surprisingly, both FPC-based mapping and hybridizations to a blot of the fingerprinting gel indicated that not all BACs fell within the contig. Rather, some BACs identified by probes A and B, both of which had weak hybridization to many BACs, did not map to the region surrounding *odc*. By pursuing those BACs that were confirmed by both hybridization and restriction fragment mapping, we were able to

place the BAC-end derived markers on the original 110 kb contig and walk outward 83 kb in one direction and 35 kb in the opposite direction. Thus, we assembled a 228 kb contig around the locus associated with the *odc* probe (Fig. 4).

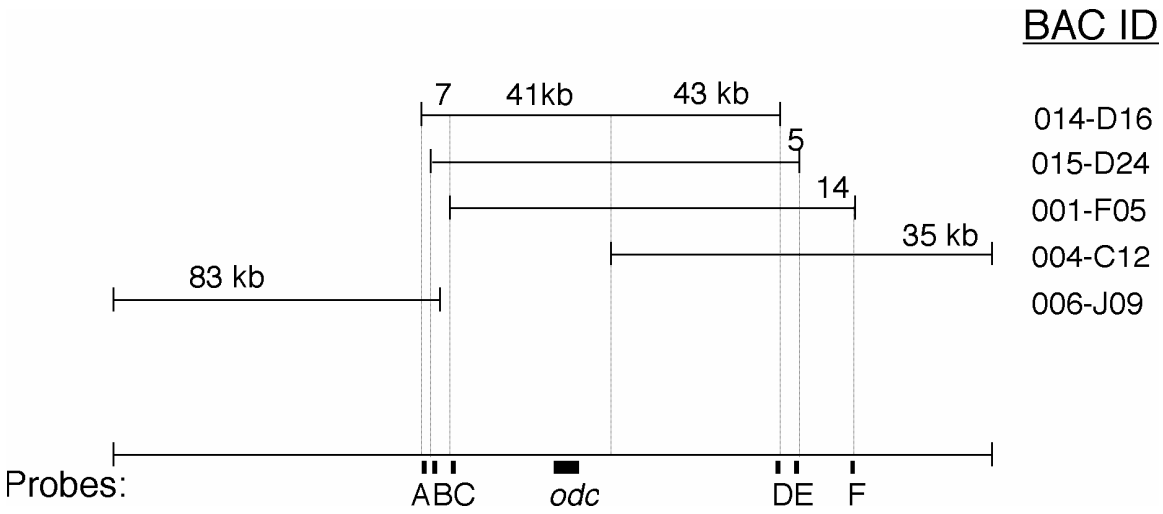

Fig. 4. A set of contiguous BAC clones associated with the *odc* probe. To the right of each BAC diagram is the clone ID number. The numbers above each BAC represent sizes (kb) of segments. BACs 015-D24, 014-D16, and 001-F05 were among the 8 BACs that hybridized to the *odc* probe (Table 1). To conduct a single walking step in each direction probes were developed based on BAC ends (indicated as A, B, C, D, E, and F). BACs 006-J09 and 004-C21 were mapped based on their hybridization to probes A through F, and are representative of the 32 BACs mapped that extend the contig to a total of at least 228 kb.

Recent studies in filamentous fungi have identified microsynteny between a BAC-sized region in *M. grisea* and the *N. crassa* genome (Hamer et al., 2001) and between sets of linked ESTs from *Blumeria graminis* to syntenic regions in *N. crassa*

and *A. fumigatus* (Pedersen et al., 2002). In this study, the availability of 32 kb of *E. festucae* sequence mapped to a 43 kb region associated with the methionine synthase marker (Table 1) presented a relatively unique opportunity to assess the degree of microsynteny between an endophyte species and some model fungi (Table 2; Fig. 5). The *E. festucae* sequence exhibited homology to 13 predicted ORFs in *N. crassa* (Table 2). We designated these 13 regions of homology ORF-A to ORF-M, however it should be noted actual elaboration of a functional protein is not confirmed for any of these putative ORFs. Furthermore, subsequent analysis of sequences with FGENESH predicted functional proteins for 11 of these 13 homologous regions. B and D were the ORFs not predicted by the program. Nonetheless, from the standpoint of analysis of microsynteny it is important to recognize the maintenance of homology between these sequences and predicted proteins in *N. crassa*.

Table 2 details the results of blastp or tblastn analysis between *E. festucae* and the genome sequences of four model fungi. Homologs to a majority of the 13 ORFs found within the 43 kb region in *E. festucae* are also found on a single contig or supercontig in *N. crassa*, *M. grisea*, and *A. nidulans*, but are widely dispersed in *C. neoformans*. The lack of synteny in *C. neoformans* is expected considering it is a basidiomycete, thus, divergence of the last common ancestor between it and *E. festucae* and the other Ascomycetes occurred approximately 550 million years ago (Berbee and Taylor, 2001). Fig. 5A depicts the map of the *E. festucae* contig assembly in which all the ORFs are anchored except for ORF-M. The ORFs are clustered into five regions of colinearity (R1-R5). The order of and distance between these regions varies across

Table 2
Identification of orthologs to 13 *E. festucae* ORFs based on BLAST searches of publically available genome sequence databases

E. festucae		*N. crassa*[b]				*M. grisea*[b]				*A. nidulans*[c]			*C. neoformans*[b]	
ID	**Functional homology**[a]	**ID**	**Contig**	**Super contig**	**e-val**	**ID**	**Contig**	**Super contig**	**e-val**	**Contig**	**Super contig**	**e-val**	**Contig**	**e-val**
A	cobalamin-independent methionine synthase [*A. nidulans*]	**NCU06512.1**	**3.372**	**28**	**0**	**MG06712.1**	**2.1250**	**12**	**0**	**1.77**	**5**	**0**	**1097**	**0**
B	hypothetical protein [*N. crassa*]	**NCU03403.1**	**3.177**	**10**	**2e-11**	**none**[d]	**2.1250**	**12**	**2e-10**	**1.51**[d]	**3**	**6e-04**[3]	**238**	**5e-06**
C	probable Gim complex component GIM3 [*N. crassa*]	**NCU03401.1**	**3.177**	**10**	**3e-15**	**MG06730.1**	**2.1250**	**12**	**3e-41**	**1.13**	**1**	**e-36**	**121**	**3e-12**
D	hypothetical protein [*N. crassa*]	**NCU03400.1**	**3.177**	**10**	**3e-04**	**none**[d]	**2.1250**	**12**	**5e-07**	**1.88**[d]	**6**	**3e-11**	**none**	**-**
E	predicted protein [*N. crassa*]	**NCU03399.1**	**3.177**	**10**	**5e-39**	**MG06729.1**	**2.1250**	**12**	**5e-73**	**1.51**[f]	**3**	**8e-04**	**60**	**3e-09**
F	HSP 70 family protein [*Schizosaccharomyces pombe*]	**NCU09485.1**	**3.606**	**76**	**e-118**	**MG06648.1**	**2.1238**	**12**	**e-136**	**1.13**	**1**	**e-100**	**1018**	**e-84**
G	endothiapepsin [*Cryphonectria parasitica*]	**NCU09484.1**	**3.606**	**76**	**5e-09**	**MG06647.1**	**2.1238**	**12**	**2e-35**	**1.51**[f]	**3**	**e-09**	**1272**	**3e-07**
H	probable nucleolar protein NOP58 [*N. crassa*]	**NCU03396.1**	**3.177**	**10**	**e-131**	**MG07008.1**	**2.1301**	**12**	**e-176**	**1.51**	**3**	**e-173**	**584**	**e-107**
I	vacuolar ATP synthase subunit D [*N. crassa*]	**NCU03395.1**	**3.177**	**10**	**e-162**	**MG07007.1**	**2.1301**	**12**	**e-172**	**1.51**	**3**	**e-123**	**1039**	**e-107**
J	ribosomal protein S15 precursor (mitochondrial) [*N. crassa*]	**NCU03394.1**	**3.177**	**10**	**2e-87**	**MG07006.1**	**2.1301**	**12**	**e-72**	**1.51**	**3**	**e-53**	**876**	**3e-04**
K	ribosome-associated protein (Rap-1) [*N. crassa*]	**NCU03393.1**	**3.177**	**10**	**8e-72**	**none**[e]	**2.1300**	**12**	**9e-48**	**1.51**	**3**	**2e-95**	**1431**	**2e-43**
L	related to LTE1 protein [*N. crassa*]	**NCU03379.1**	**3.176**	**10**	**0**	**MG02419.1**	**2.491**	**3**	**0**	**1.51**	**3**	**e-154**	**1076**	**2e-05**
M	stomatin-like protein [*Gibberella fujikuroi*]	**NCU03388.1**	**3.177**	**10**	**e-129**	**MG02410.1**	**2.488**	**3**	**e-131**	**1.51**	**3**	**7e-81**	**666**	**3e-56**

a Function listed is based on best blastx hit to a protein with a predicted function or homology given. Species corresponding to this predicted function is provided in brackets.
b Homology between *E. festucae* ORFs predicted by FGENESH and annotated proteins in *N. crassa* or *M. grisea* using blastp unless otherwise stated.
c Homology between *E. festucae* ORFs predicted by FGENESH and genome sequence data from *A. nidulans* and *C. neoformans* using tblastn unless otherwise stated.
d No ORF was predicted by FGENESH for *E. festucae* or *M. grisea* , however tblastx identified nucleotide homology in the region.
e No protein was predicted in the genome annotation, but homology was identified in the genome sequence via tblastn.
f The match given is not the highest scoring BLAST hit for that ORF but is included because this is the highest match showing synteny.

E. festucae, *N. crassa* and *M. grisea* (Fig. 5A). All of the corresponding ORFs in *M. grisea* are located on linkage group II. Interestingly, most of the small regions of colinearity are widely distributed across > 1 Mb in that fungus. By way of contrast, the majority of presumed orthologous ORFs in *N. crassa* fall within an 84 kb region which maps primarily to linkage group II. However, R3 in *N. crassa* is >2.5 Mb away from the other regions within linkage group II and R1 appears to map to linkage group III (Fig. 5A). It is interesting to note that proximity between R4 and R5 is conserved in the comparison between *E. festucae* and *N. crassa* but not between these two and *M. grisea*. Furthermore, R1 and R2 which are adjacent in *E. festucae* are within 13.2 kb in *M. grisea* but located on different chromosomes in *N. crassa*. This is surprising given the fact that *N. crassa* and *M. grisea* are more closely related to each other than either is to *E. festucae*. Fig. 5B depicts the order and orientation of these ORFs within the regions of colinearity. Within each region all intervening ORFs are depicted for *N. crassa* and *M. grisea* (as indicated by solid bars), additionally since R2, R4, and R5 are adjacent in *N. crassa* all ORFs are depicted for that 84 kb segment. Given the relatively small amount of sequence data for *E. festucae* it would be unwise to focus intensive scrutiny on the microsynteny differences within each of these regions, especially relative to the ORFs present in the other two fungal species but with no corresponding sequence available in *E. festucae*. Generally speaking it is clear that a third, more distantly related taxon *(E. festucae*) serves well to focus ORF predictions for the two more closely related fungi. By way of example, orthologs to ORFs B, D and K were not predicted in the *M. grisea* proteins release 2.1, but conservation of genomic sequence corresponding to

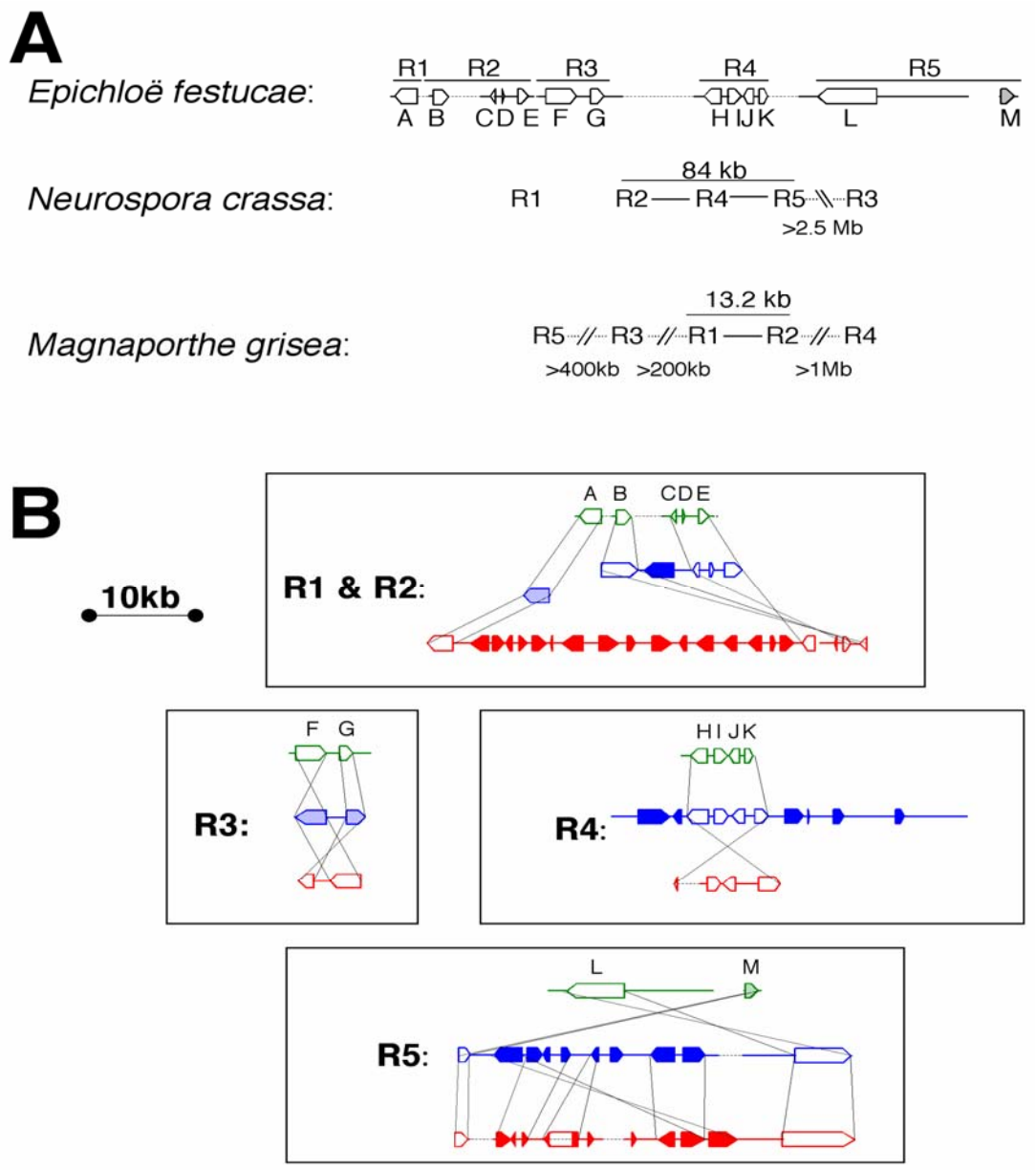

Fig. 5. Microsynteny analysis for genomic region associated with the T7 end of BAC 14-O9. **A.** An overview of the region. Thirteen ORFs predicted in the available *E. festucae* sequence (Genbank accession AY365418) are based on a variety of BLAST comparisons. Sequence contigs were placed on this map based on restriction enzyme digests, end sequencing of subclones, and anchoring with some BAC ends in the region. ORFs with striped pattern are not anchored, thus comparison of orientation is not possible. Five regions (R1-R5) were assigned in order to simplify representation of the patterns of microsynteny. Order, orientation, and degree of expansion or loss of genes varies across these regions for the three species. **B.** Comparison of maps by regions of microsynteny. Maps are color-coded (*E. festucae* = green, *N. crassa* = blue, *M. grisea* = red). Dashed lines represent gaps between contigs, thus exact distance is unknown. Solid arrows represent ORFs present in the public genome sequences not present in this *E. festucae* sequence.

these ORFs among all three taxa focused identification of these missing ORFs more than might have occurred in a pairwise comparison of *N. crassa* and *M. grisea*.

There is a multitude of uses for the *E. festucae* BAC library including: positional cloning, comparisons of genome structure between *E. festucae* and other filamentous fungi, and the planned construction of an integrated genetic and physical map. An obvious application of genome structure comparison would involve addressing the structure of the often heteroploid *Neotyphodium* endophytes (Craven et al., 2001a; Schardl et al., 1994; Tsai et al., 1994). In fact, use of the haploid *E. festucae* to elucidate the genetics of the related heteroploid agriculturally important *Neotyphodium* endophytes is directly analogous to use of plants with simple genetic systems to study more complex crop species (e.g. use of *Meticago truncatula* as a model for legume crops (Frugoli and Harris, 2001).

Since endophyte derived host fitness enhancements have frequently been linked to expression of bioactive alkaloids (Bush et al., 1997; Schardl, 2001) an obvious target for map-based cloning in endophytes would be genes associated with alkaloid expression. *E. festucae* is unique among endophytes in that it possesses both the ability to execute a sexual cycle and the ability to produce each of the four known classes of alkaloids. Furthermore, since genes for secondary metabolism tend to be clustered in fungi (Keller and Hohn, 1997) identification of markers linked tightly to alkaloid expression bodes well for marker assisted gene discovery specific to *E. festucae* fitness components like loline alkaloids (Wilkinson et al., 2000).

A BAC-based physical map would be an excellent asset for endophyte research. While high sequence coverage genomics efforts might be quite a reasonable proposition for some other filamentous fungal models (*Neurospora crassa, Magnaporthe grisea, Aspergillus nidulans*) planned genomics efforts for endophytes are more likely to provide for simple gene discovery, involving low levels of sequence coverage (1X or less).* Thus, the importance of an established physical framework for localizing genes of interest, like that provided by this BAC library resource and any subsequent BAC-based physical map is especially pronounced.

Casting an eye toward the future directions of comparative fungal genomics, it is clear that while established model species will provide basic insight into how all fungi live, a wide variety of fungal models will be necessary to elucidate the unique features of any particular life cycle. Continued development of *E. festucae* with resources such as this BAC library will assist us in assessing the features of a highly intimate plant-fungus mutualism. Furthermore, *E. festucae* possesses the ability to express both vertical transmission through the host seed and a more aggressive, horizontal transmission strategy on different tillers of the same plant. Thus, it should be possible to locate and identify genetic mechanisms responsible for transition between being a replacement pathogen (analogous in many ways to the closely related *Claviceps* spp.) vs. a benign endophytic mutualist.

---

*Since the publication of this manuscript, an *E. festucae* genome sequencing project has commenced with Christopher Schardl's group at the University of Kentucky. This sequence is expected to be released by 2007.

CHAPTER III

COMPARISON OF LOLINE ALKALOID GENE CLUSTERS ACROSS FUNGAL ENDOPHYTES: PREDICTING BOTH THE CO-REGULATORY SEQUENCE MOTIFS AND THE EVOLUTIONARY HISTORY OF THE TRAIT

Introduction

Comparative sequence analysis across intraspecifc lineages, and closely related species, provides both for identification of conserved sequence elements and for tracing the evolutionary history of a trait. The genetic basis for the expression of insecticidal loline alkaloids by fungal endophytes of grasses is the *LOL* secondary metabolite gene cluster (Wilkinson et al., 2000; Spiering et al., 2002; Spiering et al., 2005) We recently described the presence of two *LOL* gene clusters in the genome of the endophyte *N. uncinatum*. Further, we noted that expression of cDNAs for *lol* genes in the tall fescue endophyte *N. coenophialum* supported the presence of a single *LOL* cluster in that fungus (Spiering et al., 2005). In this study the sequence from five *LOL* clusters from four endophyte species (single clusters from *Epichloë festucae, Neotyphodium coenophialum, Neotyphodium* sp. PauTG-1and the two clusters from *Neotyphodium uncinatum*) provide for a powerful comparison of sequence conservation.

At least nine genes associated with lolines expression are known (Spiering et al., 2002; Spiering et al., 2005) and all are physically linked in one or more *LOL* clusters. Three genes, *lolC*, *lolD* and *lolT,* are homologous to genes that encode pyridoxal-5'-phosphate (PLP) dependent enzymes important or essential for primary metabolism (*O*-acetylhomoserine(thiol)-lyase, ornithine decarboxylase, and an aminotransferase, respectively). The predicted protein encoded by the *lolA* sequence has similarity with an aspartate kinase C-terminal domain but lacks the kinase active site (Spiering et al., 2002). Four other *lol* genes, *lolE*, *lolF*, *lolO* and *lolP*, are predicted to encode oxidation enzymes. The predicted *lolU* gene product exhibited homology to fungal proteins of unknown function. A putative tenth *lol* gene, *seqX*, is physically linked to *lolF* in one *N. uncinatum* cluster (Spiering et al., 2005). *seqX* is transcribed during lolines producing conditions but has no predicted function and might be a pseudogene (Spiering et al., 2005.) In the analysis of the two *N. uncinatum* clusters, we noted that the order and orientation of the genes was strictly conserved, with several gene pairs sharing the 5'region, perhaps indicating shared regulatory sequences between these pairs. In addition to allowing for shared regulatory elements, such an arrangement might result in selection to conserve the order and orientation of the *lol* genes during maintenance of the cluster in particular endophyte genomes. In the present comparison of five *LOL* clusters we continued to test the hypothesis of shared regulatory elements selecting for conservation of order and orientation of *lol* genes. To that end, we sought to identify and compare putative transcription factor binding sites conserved in the 5' non-coding regions of the *lol* genes.

An important aspect of the evolution of grass-endophyte associations is that almost all of the seed-transmitted asexual lineages, *Neotyphodium* spp. are the product of ancient interspecific hybridization events. It is thought that the fact that these species are aneuploid, based on possession of more than one copy of many housekeeping genes (Tsai et al., 1994; Moon et al., 2004), is evidence for their hybrid origins. Although this phenomenon was first identified in the endophytes (Tsai et al., 1994) it has subsequently been recognized in several other fungal species (Brasier, 2000; Schardl, 2001; Schardl and Craven, 2003) suggesting that the frequent involvement of *E. festucae* as an ancestor in the hybrid endophytes was perhaps due to selection for alkaloids. Furthermore, Clay and Schardl (Clay and Schardl, 2002) suggest that as the only sexual endophyte known to possess some alkaloid classes, including lolines, *E. festucae* likely contributed many of the alkaloid biosynthesis genes to the seed-transmitted *Neotyphodium* endophytes that produce these alkaloids. However, they also point out that lolines biosynthesis may represent a special case, because based on phylogenetic analyses of lolines expressing species, *E. festucae* is *not* always implicated as a contributor to their hybrid genomes. In this study we analyzed sequences from four closely related species to help elucidate the evolution of the loline alkaloid biosynthetic cluster, *LOL*. These analyses allowed us to determine the most likely origin of the *LOL* cluster(s) for three *Neotyphodium* species with different ancestries, including two species without any history of the *E. festucae* genome implicated as a progenitor (*N. uncinatum* and *Neotyphodium* sp. PauTG-1) and one with *E. festucae* previously established as an ancestor (*N. coenophialum)*.

Materials and Methods

Fungal DNA extraction, PCR, sequencing, and Southern hybridization

Fungal material for isolates used in this study (*E. festucae* 189 (CBS 102477), *E. festucae* 434 (CBS 102474), *E. festucae* 1035.30 (CBS 102475), *N. coenophialum* (ATCC90664), *N. uncinatum* (CBS 102646) and *Neotyphodium* sp. PauTG-1 (isolate e55)) was prepared as described by Blankenship et al. and DNA was isolated by the method described by Al-Samarrai and Schmid (Al-Samarrai and Schmid, 2000; Blankenship et al., 2001).

Unless otherwise indicated, PCR amplification was performed using AmpliTaq Gold DNA Polymerase in 25 μL reactions with 40 ng genomic DNA, 2 mM dNTPs, 2 mM $MgCl_2$, and 0.2 μM of each primer. Most reactions were 40 cycles with a 50°C annealing temperature and 2 min extension at 72ºC. Amplification products were purified through a Bio-Rad Spin 50 mini column and then either sequenced directly or cloned using pCR4-TOPO (Invitrogen) and sequenced. Sequencing reactions were performed as described in Chapter II (Kutil et al., 2004). Sequences were edited and assembled using Sequencher 4.1 software (Gene Codes, Ann Arbor, Mich.).

Membranes for Southern hybridizations were prepared as described previously (Kutil et al., 2004). To generate probes TOPO-cloned PCR products were digested with *Eco*RI, allowed to separate on a 1% agarose gel, then purified using the Qiaquick gel extraction kit (Qiagen). The probe DNA (25 ng) was labeled using a standard High

Prime reaction (Roche). The hybridization and membrane washing were performed as described in Chapter II (Kutil et al., 2004).

Generating sequences for LOL clusters

Mapping and sequencing LOL in E. festucae: In a previous study we bred two field isolates of *E. festucae* naturally polymorphic for loline alkaloid production (*E. festucae* 434 is Lol+; *E. festucae* 189 is Lol-). An AFLP fragment segregating 100% with lolines production in the progeny was identified (Wilkinson et al., 2000). Once cloned, the AFLP fragment was used to probe an 18X *E. festucae* BAC library constructed from a lolines producing progeny isolate (*E. festucae* 1035.30) (Kutil et al., 2004). One 100 kb BAC (14-O9) that hybridized to the cloned AFLP fragment was initially mapped and characterized using a variety of restriction enzymes, cloning resulting fragments into pBluescript, and sequencing both ends. Identification of many putative *lol* genes prompted us to sequence the entire BAC. The BAC DNA was sheared and subcloned into a shotgun library. Inserts for 1056 clones were sequenced from both directions resulting in >10X coverage of this region of the *E. festucae* genome *(*Kutil et al., 2004). In this sequence we identified the *LOL* locus, an approximately 40 kb region (at the SP6 promoter end of BAC 14-O9), present in the Lol+ parent and absent in the Lol- parent. We used cDNA data from the simultaneous study of *N. uncinatum* (Spiering et al., 2002) and FGENESH (http://www.softberry.com/berry.phtml) to identify eight putative open reading frames implicated in the biosynthesis of loline alkaloids (*lol*E, *lol*T, *lol*P, *lol*U, *lol*A, *lol*O, *lol*D, and *lolC)*. Since there were two additional features linked to *lolC* in the

LOL cluster(s) of *N. uncinatum (lolF* and *seqX*) we attempted to PCR amplify these sequences from the *E. festucae* genome. We were able to retrieve *lolF* from *E. festucae* 1035.30 genomic DNA using (primers 5' CCGTGCAGAGATGCTGAG and 5' CCCTTGTGCAACCTACC, with 45 cycles and a 55ºC Tm). Subsequently we used this *lolF* product as a probe against the *E. festucae* BAC library. No BACs in the 18X library hybridized to this sequence. For *seqX* repeated attempts at PCR amplification from the *E. festucae* genome were unsuccessful. Therefore, we performed an interspecific hybridization experiment by probing the *E. festucae* BAC library with a 300 nucleotide *seqX* PCR product amplified from *N. uncinatum* (primers SeqX Up AGGCGATACCCGAGCTTC and SeqX Low, GCCTCGATTGCGTCATAG). This probe hybridized to only one BAC (BAC 02-E8). After end sequencing this BAC and developing primers and probes associated with these end sequences, we determined that, as was the case for BAC 14-O9, one end of BAC 02-E8 was present in the Lol+ parent and absent in the Lol- parent. Thus, we believe that these two BACs flank the region in the *E. festucae* genome responsible for the Lol +/- polymorphism. Further analyses of BAC 02-E8 indicate that the *seq*X probe hybridized to a 4-kb *Xho*I fragment at the SP6 end of BAC 02-E8. Within the sequence of that fragment we identified 100 nucleotides similar to *seqX*. Sequencing in BAC 02-E8 revealed that *orfN* is the next ORF flanking *lolX* in *E. festucae* (Fig. 6). *orfN* has homology to a vesicle fusion protein *nsf1* in *N. crassa* (**NCU03387.2**). *orfN* and sequences from the T7 end of BAC 02-E8 are present in both the Lol+ and Lol- *E. festucae* parental isolates (Fig. 6A).

Sequencing LOL clusters from additional Neotyphodium species: The sequences of two copies of the *LOL* cluster (*LOL*1 and *LOL*2) from *Neotyphodium uncinatum* are well characterized (Spiering et al., 2005). In this study, to generate sequence for additional *LOL* clusters from *N. coenophialum* and *Neotyphodium* sp. PauTG-1, primers were designed based on both the *N. uncinatum* sequences and the *E. festucae* sequence (Table 3). To verify polymorphisms, most regions were sequenced in triplicate from independent PCR reactions.

Analysis of microsynteny and homogeneity among clusters

Assessing order, orientation and intron-exon structure of nine *lol* genes in each of the five clusters (**AY723749.1**, **AY723750.1**, **EF012265** , **EF012266** , **EF012267**, **EF012268**, **EF012269**, **EF015399**, **EF015400**) involved alignment using MultAlin (Corpet, 1988); http://prodes.toulouse.inra.fr/multalin/multalin.html) and prediction of coding sequences based on previously identified cDNAs (Spiering et al., 2002; Spiering et al., 2005) and FGENESH (Softberry Inc.; http://www.softberry.com/berry.phtml) trained for the closely related ascomycete *Neurospora crassa*.

To assess the degree of homogeneity among the clusters, regions spanning most or all of nine *lol* genes for each cluster were compared. In PAUP 4.0 independent maximum parsimony trees were generated for each coding and non-coding region. A partition homogeneity test was performed to determine whether or not combining the data was justified for a single tree. In all maximum parsimony analyses heuristic search was performed with stepwise addition set to random for 1000 replications with new

Table 3
Primers used for PCR and sequencing

Primer Name [a, b]	Sequence
lolE to lolP Up	CAAGTCTGCGCTCCCAC
lolE to lolP Low	CTCTCTTTGGGTCTGTCC
lolE to lolP F1	GGCATCTACGGCGATCA
lolE to lolP F2	CGAAGATATCCGTTGGCTTGG
lolE to lolP F3	GGAGGCTTTCTGCATGG
lolE to lolP F4	ATCAACTACTTACGGTG
lolE to lolP F5	CTTCGTCTCTGATTGCCAC
lolE to lolP F6	ATATGCGAAGAGATACTC
lolE to lolP R1	CCGAAATTTTGCGTCACG
lolE to lolP R2	CGGCAAGGATGGGATAG
lolE to lolP R3	GGCGAGGTGGTCGAGGTC
lolE to lolP R4	GCAGGAACATTGCTCTC
lolP to lolD-1 Up [c]	GTATTGACCCCAGGTTATGG
lolP to lolD-1 Low	GGATAGTCTGGGAGAGTAAAGC
lolP to lolD-1 F1	CGCAATGCAATGTAACATT
lolP to lolD-1 R1	TAATATAACAAAGGGCATT
lolP to lolD-2 Up	GGTAAAGTCGCCCAAATAGG
lolP to lolD-2 Low	GCACAAGACACTCCTTGTG
lolP to lolD-2 F1	CGAGTATTGTCCGCGTC
lolP to lolD-2 R1	GAAGTGGACGCAGGTTCGC
lolP to lolD-3 Up	GAGATTTTGTGGGTGGGCTG
lolP to lolD-3 Low	CATGTTCTGGAGAGATTGC
lolP to lolD-3 F1	CCAATAATAGAGACCGTGTC
lolP to lolD-2 R1	CTTGTTTGTTACCGTCATC
lolP to lolD-4 Up	TGGAGCAGCATCTCCCG
lolP to lolD-4 Low	CACCACCTTCAATGGCTTCAC
lolP to lolD-4 F1	TGCCAGTCGGTCAGAAAC
lolP to lolD-4 R1	GAGAACGGTAAAAATGACG
NP lolC to lolD-1 Up	GTATGTACATGTAGGTATCAGTCTAGCGCG
NP lolC to lolD-1 Low	CGACGTCGTTCGTAAGTGGGCGG
NP lolC to lolD-2 Up	GAAATGGGGGGGAATTGAAGACGTACC
NP lolC to lolD-2 Low	GGATATCAAGGCCGACTTCGAGCAGG
lolC Up	CGGGAACCAAGATGTTCC
lolC Low	ACGACAAGCTGAAGGTCC
lolC F1	CATCGACGACAAGACCAAG
lolC F2	CAAGTGGATCGGCGG
lolM Up	CCCTCATTGACAACTCGG
lolM Low	CGTCAATACTTATAGAGGTCTC
NC lolF Up	CAGTATGCGCGGTAAGAGGG
NC lolF Low	GGGTGTTTGTCAATACCAGAG
NP lolF Up	CCAGGAATGTCTCGATTCCGC
NP lolF Low	GAGTTGACCAGAAGTGTAGG
lolF to lolM-1 Up	GGAGGCGGCTCTTCTGTGAAGCCACTTTCTGTCG
lolF to lolM-1 Low	TTTTGATTCAAGTCATCAACAACCTCCGAGT
lolF to lolM-1 F1	GCCACATTAAAATAGGTGGG
lolF to lolM-1 R1	CGAGAGAGCTCGAGAGAAG

[a] 'Up' and 'Low' indicate primers used for PCR amplification.
'F1', 'R1', etc. indicate primers used to sequence the previously amplified fragment.

[b] 'NC' or 'NP' at the beginning of the primer name indicates that the primer was used only in *N. coenophialum* or *Neotyphodium* sp. PauTG-1 respectively. All other primers were used in both species.

[c] The lolD to lolC region was amplified and sequenced via four smaller PCR regions.

number seeds for each independent analysis. Ultimately to generate a single tree the available predicted coding sequences for the nine *lol* genes in each of five clusters were concatenated using MacClade (Sinauer). Next, a maximum parsimony tree was inferred in PAUP 4.0. Bootstrap values were also calculated in PAUP 4.0 using the heuristic search and 1000 replications.

Identification of promoter region motifs

To test *in silico* for potential co-regulation of the *lol* genes we searched for common motif sequences associated with seven *lol* genes. Putative promoter regions, ranging from 400-2000 base pairs, from *lolE*, *lolT*, *lolP*, *lolU*, *lolA*, *lolO*, and *lolD* (Fig. 8) were screened with RepeatMasker (http://www.repeatmasker.org/) to eliminate interspersed repeats and low complexity DNA. The sequences were then aligned using ClustalW (Thompson et al., 1994) with a Gap Opening Penalty of 15.0 and a Gap Extension Penalty of 0.3. Insertions and deletions were removed using Gblocks (Castresana, 2000) using the default settings. Next, PhyloCon (Phylogenetic Consensus; Wang and Stormo, 2003) was used to generate sequence motifs. Phylocon is an algorithm designed to capitalize on both conservation among orthologous genes and the co-regulation among different genes to identify putative regulatory motifs. Because the pairs of genes *lolE* and *lolT*, *lolP* and *lolU*, *lolA* and *lolO* each share a common upstream region, we defined four orthologous groups, each with sequence from all five clusters (*E. festucae*, *N. uncinatum* clusters *LOL*1 and *LOL*2, *N. coenophialum*, and *Neotyphodium* sp. PauTG-1). With the repeats, low complexity DNA, and

insertions/deletions removed, sequences within a given orthologous region from each of the five clusters were quite conserved (at least 85% DNA sequence similarity). PhyloCon created profiles of local multiple alignments in each orthologous region and then compared the profiles to each other to determine if significant similarities exist among regions. Stringencies ranging from 0.5 to 2.0 were tested for each of the PhyloCon analyses.

We also analyzed two sets of control sequences with PhyloCon. The first set included promoter regions for four genes from five publicly available fungal genomes as well as *E. festucae*. The genes used are orthologous to a set of defined genes from the T7 end of BAC clone 14-O9 of *E. festucae* (Kutil et al., 2004). There is no reason to expect *a priori* that this group of genes should be co-regulated. While they are physically linked to each other in *E. festucae* (and to a large extent in at least two other fungal species; (Kutil et al., 2004), the genes are not involved in related metabolic functions. These control genes come from *Neurospora crassa, Aspergillus nidulans, Magnaporthe grisea,* and *Fusarium graminearum*. The genes are a HSP 70 family protein (**AAQ73633**, **FG10819.1**, **MG06648.4**, **NCU09485.1**, **AN0847.2**), a probable nucleolar protein NOP58 (**AAQ73635.1**, **FG10908.1**, **MG07008.4**, **NCU03396.1**, **AN3167.2**), a Vacuolar ATP synthase subunit D (**AAQ73636, FG10907.1, MG07007.4, NCU03395.1, AN3168.2**), and ribosome-associated protein S15 precursor (**AAQ73637.1, FG10906.1, MG07006.4, NCU03394.1, AN3173.2**). The second control set is the downstream shared regions of the *LOL* gene cluster (*lolP* to *lolT*, *lolU* to *lolA* and *lolO* to *lolD*). In this control set the comparison consists of the same set of closely related endophyte

species as for the experimental comparison; however, there is no reason to expect co-regulatory elements downstream from these *lol* genes. Both sets of controls were filtered with Repeat Masker. Large-scale alignment of the control sequences with ClustalW was not always successful. In these cases, PhyloCon was used to generate motif profiles based on short local alignments.

Finally, motifs from putative promoter regions and from control sequences generated by PhyloCon were compared against the TRANSFAC 6.0 motif library using the classification "fungi" and using a cutoff score of 65 (Wingender et al., 2001). No gap penalty was specified.

Results

Characteristics of the LOL cluster in E. festucae

Fig. 6 depicts the hypothesized arrangement of BACs spanning *LOL* in *E. festucae*. Included in the diagram are the locations of *lol* gene sequences, additional flanking open reading frames and also numbers indicating representation of particular regions and markers in the BAC library. *lolF* is not located on a BAC, however we placed it in the uncloned region between the non-overlapping BACs based on its location in other *LOL* clusters and its presence in the *E. festucae* Lol+ isolates and absence in *E. festucae* Lol- (Fig. 7)

Although *lolF* is not represented in the BAC library, *lolF* sequence was amplified by PCR from the Lol+ endophytes (*E. festucae* isolate 434 and progeny 1035.30,

N. coenophialum and *Neotyphodium* sp. PauTG-1) and did not amplify from the Lol- *E. festucae* parent 189. Furthermore, Southern hybridization with a *lolF* probe against genomic DNA from Lol+ and Lol- *E. festucae* strains clearly showed that *lolF* is present only in the lolines producer (Fig. 7). As a control, the same membrane was hybridized with a probe from the primary metabolism ornithine decarboxylase (*odc*) gene. As expected the *odc* probe hybridized to both Lol+ and Lol- strains (Fig. 7).

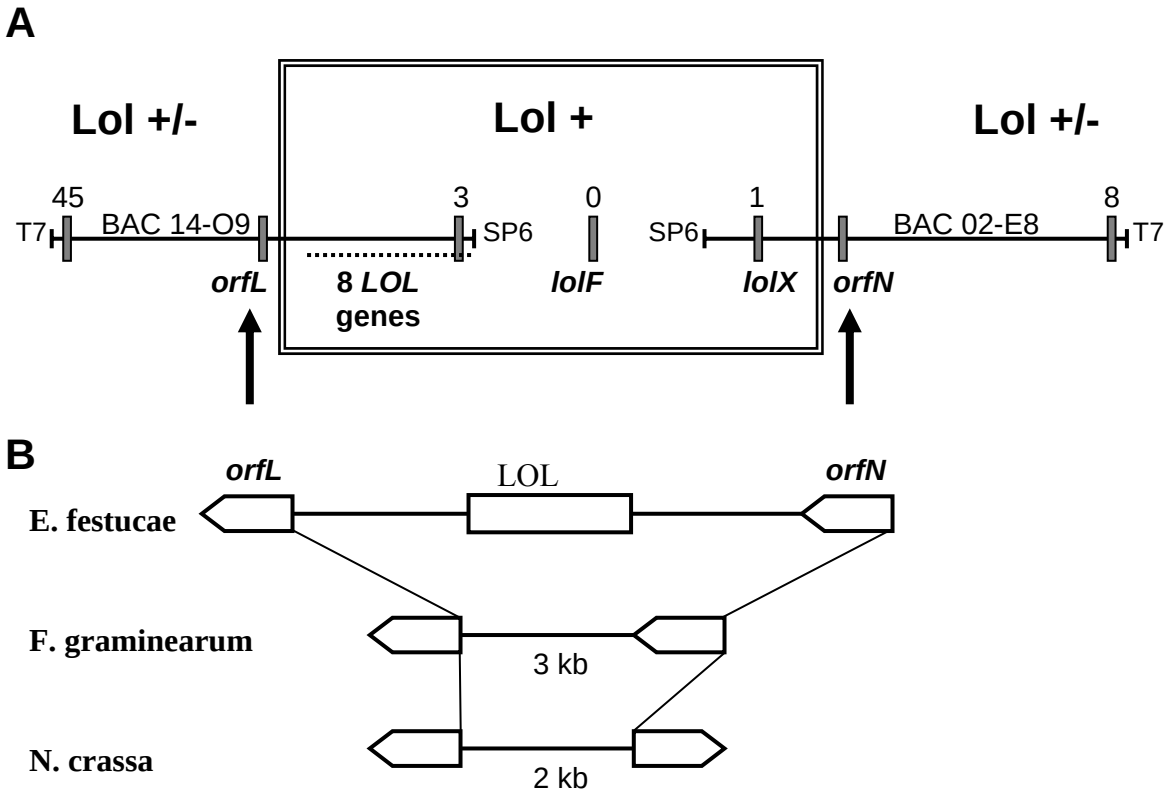

Fig. 6. Physical location of *LOL* in the *E. festucae* genome. A. Map of *E. festucae* 1035.30 BACs associated with the *LOL* gene cluster. Shaded rectangles indicate PCR markers (based on sequence from genes or BAC-ends) used to identify the limits of the polymorphism between the Lol+ and Lol- parents. The number above each marker indicates the representation of that marker in the BAC library. The boxed region including approximately 40kb from BAC 14-O9 and 5 kb from the SP6 end of BAC 02-E8 is unique to the Lol+ parent. *lolF* is present in the genome of the Lol+ parent, but is not represented in the BAC library. B. A schematic representation of the genomic arrangement of orthologs from two closely related filamentous fungi for two genes flanking *LOL* in *E. festucae*.

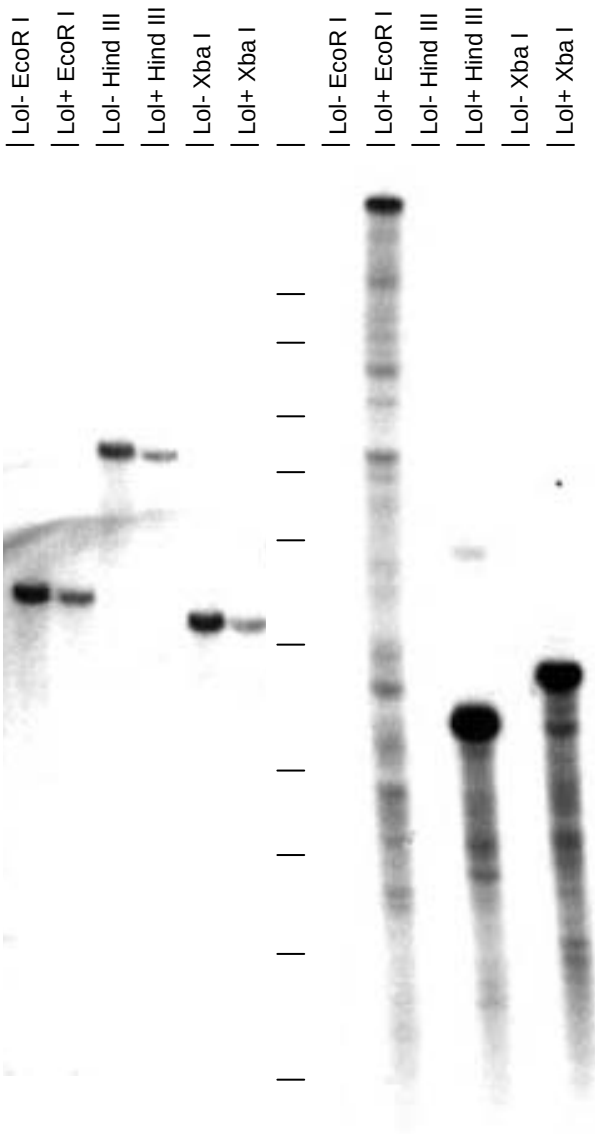

Fig. 7. Two Southern blots of a primary and a secondary metabolism gene in *E. festucae*. On the left, digested genomic DNA from lolines producing 1035.30 (Lol+) and non-producing 189 (Lol-) *E. festucae* strains is probed with the primary metabolism *odc*. On the right, the same membrane is probed with a cloned *lolF* fragment from the Lol+ *E. festucae* strain.

The next feature in *N. uncinatum LOL*2 is *seqX* (Spiering et al., 2005). This sequence is also of interest because it is an EST differentially expressed under loline producing conditions for *N. uncinatum*. Based on our comparative analysis, we have renamed *seqX lolX* (Fig. 8). No other potential ORFs were identified by FGENESH or by BLAST comparisons in the 1.8 kb region between *lolX* and the SP6 end of BAC 02-E8. *orfN* is the next ORF flanking *lolX* in *E. festucae* (Fig. 6). *orfN* has similarity to a vesicle fusion protein gene *nsf1* in *N. crassa* (**NCU03387.2**). *orfN* and sequences from the T7 end of BAC 02-E8 are present in both the Lol+ and Lol- *E. festucae* parental isolates (Fig. 6A).

Our best approximation of how these genes, features and markers map relative to each other in the *E. festucae* genome is represented in Fig. 6A. In support of this arrangement the degree of representation of particular sequences in the BAC library follows a logical trend. Starting on the left of the figure a marker at the T7 end of BAC 14-O9 was present in the library in 45 BAC clones. Once again, the SP6 end of BAC 14-O9 is only present in 3 BAC clones. *lolF* is not present in the BAC library. *lolX* maps 1.8kb away from the SP6 end of BAC 02-E8 and this region is present only in this one BAC. The T7 end sequence of BAC 02-E8 is present in 8 BAC clones. Taken together, these data and the data for presence and absence of each of these features in the Lol+ and Lol- *E. festucae* strains support the hypothesis that these BAC clones are adjacent to each other in the genome but the region between them was not clonable during the construction of our 18X BAC library (Kutil et al., 2004).

orfL is a marker between the T7 region of BAC 14-O9 and *LOL*. This sequence has similarity to an LTE1 related protein in *N. crassa* (**NCU03379.1**) and was previously used in an analysis of microsynteny across filamentous fungi (AY365418.1; Kutil et al., 2004). Of the predicted coding regions characterized and anchored in that microsynteny analysis, *orfL* maps closest to *LOL* and can be detected by PCR from Lol+ and Lol- *E. festucae* isolates. Thus *orfL* and *orfN* flank either side of the *LOL* gene cluster (indicated by arrows in Fig. 6). Their orthologues are also closely linked in the published genome assemblies from *N. crassa* and *F. graminearum* (Fig. 6B) where they are separated by less than 3 kb and have no intervening genes. However, the orientation of *orfN* is reversed in *N. crassa* when compared to that of *F. graminearum* and *E. festucae* (Fig. 6B).

Sequences with similarity to pol elements mapping near LOL in E. festucae

Between the two gene-rich ends of the *E. festucae* BAC insert is an approximately 25 kb region containing at least five segments with similarity to elements that encode reverse transcriptase-like polyproteins (*pol*). Evidence of an additional element with similarity to a *pol* protein is found in the 19kb non-coding gap between *lolD* and *lolC* (Fig. 6). These six features range from 0.8kb to almost 6 kb in length, with BLAST expect values between E-12 and E-50. It is not clear that any of the elements are currently functional as a retrotransposons because most of the sequences are highly corrupted lacking definitive open reading frames and other features necessary to be functional. One of these putative elements appears to be duplicated, but the other four

appear to be unrelated. One element has homology to the Tad1-1 retroelement in *N. crassa*, and one is similar to a gene encoding a *pol* protein found in *A. nidulans* (**AN5254.2**) and at least 9 genes in *M. grisea* (**MG0168.4, MG05517.4, MG03445.4, MG10125.4, MG05786.4, MG04275.4, MG02322.4, MG02356.4, MG05548.4,** and **MG03954.4)**. None of the elements here are related to the AT-rich retrotransposons identified flanking the recently characterized lolitrem alkaloid gene cluster of the endophyte *Neotyphodium lolii* (Young et al., 2005).

Conserved microsynteny across all 5 LOL clusters

Primers based on the three known *LOL* cluster sequences from *E. festucae* and *N. uncinatum* (Table 3) were used to amplify the cluster from two additional lolines producers *N. coenophialum* and *Neotyphodium* sp. PauTG-1 (Fig. 8). Approximately 16.5 kb of contiguous sequence spanning *lolE*, *lolT*, *lolP*, *lolU*, *lolA*, *lolO*, *lolD* and *lolC* was amplified from *N. coenophialum*. A 4.3 kb contig spanning *lolF* and *lolX* also amplified in that species, however, it has not been anchored to the larger *LOL* contig. For *Neotyphodium* sp. PauTG-1, 15.6 kb of contiguous sequence spans *lolE*, *lolT*, *lolP*, *lolU*, *lolA*, *lolO*, *lolD* and all except for 78 bases from the 5' end of *lolC*. Also, we amplified and sequenced a 4.3 kb contig containing *lolF* and *lolX* in this species, which remains unanchored. For both *N. coenophialum* isolate e19 and *Neotyphodium* sp. PauTG-1 isolate e55, multiple LA PCR attempts have failed to amplify the region between *lolC* and *lolF*. In each of the five clusters, every gene anchored thus far has absolutely conserved gene order and orientation (Fig. 8).

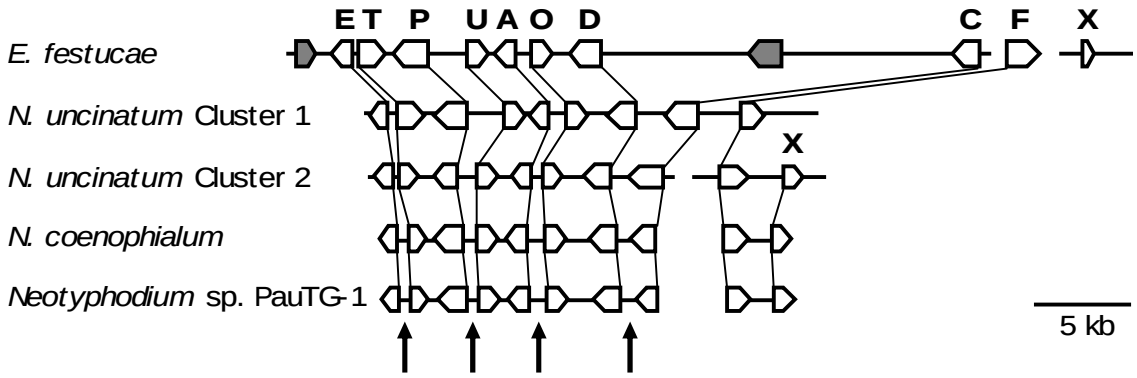

Fig. 8. Comparative maps of the *LOL* gene cluster(s) from four endophyte species. The shaded boxes in the *E. festucae* map indicate regions with homology to *pol* proteins. Disconnected contig lines indicate genes that are sequenced but not anchored to other *LOL* genes. The contig of sequence known for this region in *E. festucae* continues beyond the boundaries shown in this figure. Arrows along the bottom of the diagram indicate locations of putative promoter regions analyzed in the motif analysis.

Intron conservation across LOL clusters

Eleven intron sites were predicted in nine *lol* genes identified in *E. festucae*. No introns are predicted in *lolX* based on comparison to cDNA sequences from *N. uncinatum* (Spiering et al., 2005). Thus, sequence and positions of a total of 55 introns were compared from the five clusters. Each *lol* gene had precisely conserved intron positions across each of the five clusters. Each intron sequence aligns readily with the orthologous intron across the five clusters, but there is no conservation between intron sequences at different positions, thus there is no evidence of intron sliding (Stoltzfus et al., 1997)

Relatedness of LOL clusters

Maximum parsimony trees were generated for each coding and noncoding region individually. In all analyses both coding and non-coding regions yielded simliar results (data not shown). Maximum parsimony analysis of each coding region for seven genes (*lolE*, *lolT*, *lolP*, *lolU*, *lolA*, *lolO* and *lolD*) and the partial sequence for two genes (*lolC* and *lolF*; Fig 7) yielded identical tree topologies. In fact, a partition homogeneity test revealed no significant difference among trees (p =1.0). Thus, the coding sequences were combined to generate a single tree (Fig. 9). Of the 9695 bp in the analysis 9178 were constant, 324 were variable but uninformative, and 193 were parsimony informative.

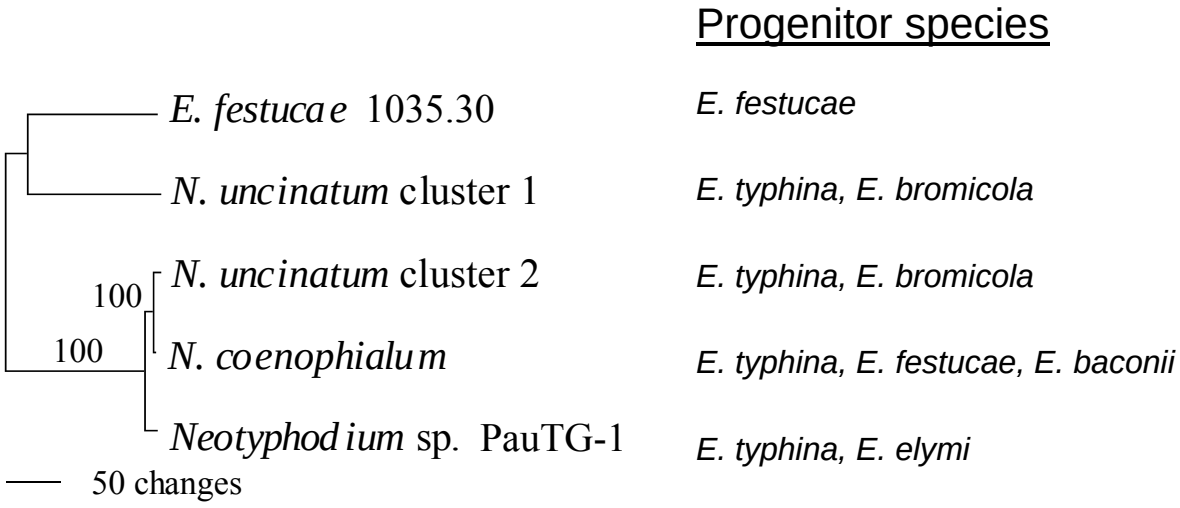

Fig. 9. Maximum parsimony tree of the *lol* coding regions. Coding sequences were concatenated for *lolE*, *lolT*, *lolP*, *lolU*, *lolA*, *lolO*, *lolD*, *lolC* and *lolF* from five copies of the *LOL* cluster in four endophyte species. Bootstrap values greater than 70 are indicated above those nodes. The progenitor species implicated in the interspecific hybridizations for each (Tsai et al., 1994; Moon et al., 2004) are listed to the right of the tree.

The *E. festucae LOL* sequence is most closely related to the sequence for *N. uncinatum LOL1*. *N. uncinatum LOL2* is most closely related to the *LOL* sequences from *N. coenophialum* and *Neotyphodium* sp. PauTG-1. Fig. 9 indicates the ancestral genomes involved in the interspecific hybridizations resulting in the *Neotyphodium* species in this study.

Conserved DNA motifs

We wanted to locate potential regulatory motifs both 1) conserved among the five *LOL* clusters, and 2) shared across promoter regions associated with different *lol* genes. Once again, three pairs of *lol* genes share upstream regions, thus, the four regions analyzed include promoters for seven *lol* genes (Fig. 8). Four different motifs with p-values of less than 1×10^{-50} (Fig. 10) resulted from this analysis. All four of these motifs were present in each of the four promoter regions analyzed (upstream from 7 *lol* genes) and in all 5 species. The PhyloCon generated consensus sequences of each of the conserved motifs in Fig. 10 were compared against the TRANSFAC database to identify the most similar fungal transcription factor binding sites. Each of the four motifs exhibited similarity to an independent binding site present in the database (ADR1, NIT2, STRE and PHO4) with a TRANSFAC score value of 65 or higher. (Fig. 10). As expected, the PhyloCon at high stringency levels (STD 2.0, 1.5, 1.0) failed to identify any conserved motifs among our control set of promoter regions (from the four genes at the T7 end of BAC 14-O9 in *E. festucae* and their orthologous promoter sequences in

	PhyloCon prediction	Closest TRANSFAC motif	Location within each block

Motif I

PHO4 score=6

Motif I	E to T	P to U	A to O	Lol D
E. Festucae	-165	6098	-9672	-13562
N. uncinatum Cluster 1	-166	5792	-9340	-13345
N. uncinatum Cluster 2	-163	4505	-7853	-11686
N. coenophylum	-163	4975	-8323	-12156
Neotyphodium sp. PauTG-1	-163	4972	-8319	-12153

Motif II

ADR1 score=7

Motif II	E to T	P to U	A to O	Lol D
E. Festucae	-254	4318	8838	-13393
N. uncinatum Cluster 1	-255	4246	8544	-13184
N. uncinatum Cluster 2	-253	3786	7235	-11517
N. coenophylum	-253	4255	7705	-11987
Neotyphodium sp. PauTG-1	-253	4253	7703	-11984

Motif III

NIT2 score=9

Motif III	E to T	P to U	A to O	Lol D
E. Festucae	361	-5492	9530	-13408
N. uncinatum 1	363	-5039	9332	-13191
N. uncinatum 2	360	-4615	7516	-11532
N. coenophylum	360	-4669	7986	-12002
Neotyphodium sp. PauTG-1	360	-4667	7982	-11999

Motif IV

STRE score=6

Motif IV	E to T	P to U	A to O	Lol D
E. Festucae	443	-5453	9357	-13385
N. uncinatum Cluster 1	445	-5000	9067	-13176
N. uncinatum Cluster 2	442	-4161	7347	-11509
N. coenophylum	442	-4630	7817	-11979
Neotyphodium sp. PauTG-1	442	-4628	7814	-11976

Control

none

Fig. 10. PhyloCon-generated logo plots for putative regulatory motifs conserved across the four promoter regions analyzed in all five *LOL* clusters. The consensus for each motif was compared against the TRANSFAC database. The TRANSFAC entry most similar to each putative motif is given, along with its score. The position of each motif is given relative to the *lolE* start site in the *LOL* cluster indicated. Negative position values indicate the site is present on the reverse complement orientation.

five model fungi). At a lower stringency level (STD 0.5), the analysis produced a motif of low complexity DNA randomly positioned among the sequences. The p-value for this motif was greater than 1 x 10^{-50} (Fig. 10); therefore, this value was used as a cutoff p-value for the dataset of *lol* promoters. Analysis of the alternative control dataset, the downstream regions of *lol* genes, also failed to identify any significantly conserved motifs. The most significant motif identified during the analysis of these control sequences was a string of A's with a p-value of 1 x 10^{-50} (Fig. 10).

Discussion

LOL was originally identified as a single genetic locus segregating in a cross of *E. festucae* isolates from two different natural host species (Wilkinson et al., 2000). Recent work has identified the individual genes clustered at the *LOL* loci of *N. uncinatum* (Spiering et al., 2002; Spiering et al., 2005). In those studies we also demonstrated that most of the genes were co-expressed under loline inducing conditions. Here we characterized the polymorphism between the *E. festucae* parental isolates used to identify the locus. We mapped and sequenced *LOL* in *E. festucae* and also sequenced *LOL* from the tall fescue endophyte *N. coenophialum* and from *Neotyphodium* sp. PauTG-1. We coupled the power of phylogenetic approaches with other forms of sequence analysis in order to infer putative functional elements, as well as the evolutionary history of the trait.

The polymorphism between the maternal Lol- *E. festucae* isolate 189 from red fescue (*Festuca rubra*) and the paternal Lol+ *E. festucae* isolate 434 from giant fescue

(*Festuca gigantean)* is an absense of the *LOL* locus in 189 rather than some sort of loss of function mutation with the gene cluster still present. All attempts to PCR amplify DNA segments within *LOL,* or hybridize genomic blots with DNA segments within *LOL* (Fig. 7), revealed presence of the sequence in isolate 434 (or a Lol+ progeny isolate 1035.30) and absence in isolate 189. The flanking regions of *LOL* are defined by the DNA segments that are present in both parents. Thus, we hypothesize that *E. festucae* isolate 189 lacks a segment of DNA that spans the *LOL* gene cluster and is at least 50 kb. However, since we do not have a BAC contig that spans the cluster in *E. festucae* it is impossible to definitively estimate the size of the polymorphism.

Clearly the most striking difference in the comparison of maps of the five clusters (Fig. 8) is the 19kb region separating *lolC* and *lolD* in *E. festucae.* In each of the other four clusters the corresponding location is no greater than 1.7 kb. The current hypothesis is that the ancestral state is the smaller region present in all the clusters except *E. festucae*. Accumulation of additional sequence between *lolC* and *lolD* in *E. festucae* is interesting in light of the presence of degenerate sequences similar to those that encode reverse transcriptases in *E. festucae*. Six such features in the *LOL* region of *E. festucae*, including one in the 19kb region between these two genes, likely indicates that this particular cluster has experienced a history of exposure to transposable elements. Interestingly, grass-endophyte associations with a lolines expressing *E. festucae* isolate tend to produce only 0.1% dry weight loline alkaloids, while grasses infected with any of the other *Neotyphodium* species analyzed produce 1-2% dry weight total lolines (Bush et al., 1997; Siegel et al., 1990; Wilkinson et al., 2000). Perhaps

disruption of the cluster by one or more retrotransposons has affected the level or efficiency of *lol* gene expression in *E. festucae*.

All *lol* genes anchored to the map for each species are in a strictly conserved order and orientation (Fig. 8). Thus, during the divergence of the different copies of the locus there does not seem to be any history of rearrangements. In addition, all genes within each cluster reflect the same relative pattern of divergence, as indicated by a non-significant partition homogeneity test. Since each cluster has evolved as a single sequence we can clearly infer the ancestry for each *LOL* cluster in each species. Perhaps the most striking outcome of this comparative analysis (Fig. 8) is that three *Neotyphodium* species with a history of *E. typhina* as an ancestor possess nearly identical copies of *LOL* (*N. uncinatum, N. coenophialum, Neotyphodium* sp. PauTG-1; Fig. 8). The simplest explanation for this pattern is that the *E. typhina* ancestor(s) to these species possessed the *LOL* cluster. This is most remarkable in light of the fact that there are no known isolates of the extant species *E. typhina* that express loline alkaloids (Leuchtmann et al., 2000). Furthermore, attempts to amplify some *lol* genes from *E. typhina* isolates did not result in a product, while all loline expressing endophytes successfully amplified with the same primers (Spiering et al., 2002; Spiering et al., 2005). The tall fescue endophyte *N. coenophialum* which we now believe has an *E. typhina* version of *LOL* also has *E. festucae* in its ancestry. Thus, prior to this characterization of the cluster from *N. coenophialum* the most reasonable expectation would have been that the cluster originated from the *E. festucae* ancestor. The one cluster that was more similar to *E. festucae* was *N. uncinatum* Cluster 1. Since the other

ancestor of *N. uncinatum, E. bromicola* is more closely related to *E. festucae* than *E. typhina*, this fits quite well with our favored hypothesis that the presence of two clusters in that genome is the product of interspecific hybridization rather than a more recent duplication event (Spiering et al., 2005).

There may be a selective advantage to the order and orientation of the genes in the cluster based on shared co-regulatory elements between genes. Our analysis of putative promoter sequences certainly supports conservation of four motifs in all the blocks within a cluster and across all clusters. Furthermore, for any given promoter the order of the motifs within the promoter was identical and the exact position was quite similar across the clusters (Fig. 10). Each of the motifs identified using PHYLOCON had a best match to a fungal transcription factor in the TRANSFAC database: PHO4, ADR1, NIT2, and STRE (Fig. 10). PHO4 is a basic helix-loop-helix protein in *S. cerevisia*e that positively regulates an acid phosphatase gene under low phosphate levels (Fisher and Goding, 1992). ADR1 is a transcription factor in *S. cerevisiae* that contains two C2H2 zinc fingers (Cheng et al., 1994). NIT2 is the major nitrogen regulator in *N. crassa* and is a C2H2 zinc finger DNA binding transcription factor (Fu and Marzluf, 1990). The *S. cerevisiae* stress response element STRE is involved in the induction of stress resistance in yeast and other organisms (Schuller et al., 1994). Since several of the transcription factor matches are associated with stress tolerance, it is tempting to speculate that the lolines may have evolved to respond to these specific stress signals. However, support for those sorts of hypotheses will require direct manipulation of those signals and their putative binding sites. Coupling the conservation of these motifs in all

the *lol* gene promoters examined with the co-expression of most *lol* genes demonstrated in previous studies (Spiering et al., 2002; Spiering et al., 2005) supports the hypothesis that the genes are co-regulated transcriptionally.

In summary this analysis of five loline alkaloid gene clusters from four endophyte species has revealed conserved sequence motifs which may serve as binding sites for a potential co-regulator as well as insights to the evolution of *LOL* clusters through three ancestral lineages of *LOL* clusters. The patterns of relatedness among clusters in these lineages support the hypothesis that *E. typhina* ancestor(s) possessed the cluster, though the trait has been lost in current day isolates. The same seems to be true for *E. bromicola* (Leuchtmann et al., 2000). Taken together with the presence vs. absence polymorphism in the *E. festucae* isolates studied, our data indicate a pattern of loss rather than gain of the trait over time.

CHAPTER IV

EVOLUTION OF TWO SECONDARY METABOLISM GENES VIA RECRUITMENT AND NEOFUNCTIONALIZATION OF PRIMARY METABOLISM GENES FROM ANCIENT SUPERFAMILIES

Introduction

There are multiple mechanisms by which a secondary metabolite gene cluster may arise in a fungal lineage (Chapter I; Walton, 2000; Zhang et al., 2004). Walton (2000) favored horizontal gene transfer (HGT) from another organism as the most likely mechanism for evolution of fungal secondary metabolite gene clusters. If lineages that aquire and express the new cluster are more fit (at least in some environments), the trait would persist. Should any lineage aquire less than the entire cluster there is no reason to expect it would have increased fitness. This mechanism is attractive because so long as HGT occurs at a reasonable frequency it would select for both formation and maintenance of the cluster. In fact, since the degree to which the cluster evolves to increase its similarity to a selfish genetic element contributes greatly to its degree of spread and persistence this is considered the selfish cluster hypothesis (Walton, 2000).

An alternative to the selfish cluster hypothesis that also favors both formation and maintenance of the cluster involves the birth and recruitment of new genes within the genome of the organism. Gene birth is often the result of a duplication event, which results in multiple copies of one or more genes in a single genome. Reduced

evolutionary constraints on the duplicated copies allows for more rapid evolution of one copy of a gene to a new function (neofunctionalization). Eventually, selective pressures may cause linkage disequilibrium among a group of co-adapted alleles. (Fisher 1930; Lawrence and Roth, 1996; Zhang et al., 2004). Under strong enough selection to link the loci, over time, it is expected the genes might become clustered where they are more likely to be maintained and inherited as a unit. Since Fisher (1930) proposed that selection would act to cause linkage among favored allele combinations, this model for clustering is referred to as the Fisher Model (Lawrence and Roth, 1996; Zhang et al., 2004).

Loline alkaloids are 1-amino pyrrolizidines with an oxygen bridge. Epichloë and Neotyphodium species are the only species known to express loline alkaloids. Bacteria produce polyamines, which were originally expected to be precursors to lolines (Bush et al., 1993), however, no pyrrolizidine alkaloids have been identified in bacteria. Pyrrolizidine alkaloids are most frequently found in particular plant families, yet their production is sporatic with respect to phylogeny. Interestingly, there is strong support for the recurrent recruitment of a primary metabolism enzyme (deoxyhypusine synthase DHS) for the enzyme (homospermidine synthase HSS) involved in the initial step in the biochemical pathways of pyrrolizidines in distinct plant lineages (Nurhayati and Ober, 2005). The pyrrolizidines expressed in plant species are expected to have a specified biochemical pathway associated with their production (Faulkner et al., 2006), thus there is no reason to expect that loline alkaloid expression evolved in one kingdom and then was transferred to another (e.g. fungi to plants, or plants to fungi).

Therefore, our favored hypothesis for the evolution of the loline alkaloid gene cluster is that *LOL* likely originated within an ancestor to the endophytes due to mechanisms similar to those in the Fisher model instead of the selfish cluster model. The hallmarks of HGT from diverse organisms include a variation in genetic architecture including intron structure, G/C content or phylogenies inconsistent with the host genome (Rosewich and Kistler, 2000). Initial analyses of homology and intron structure indicates that the *lol* genes contain introns typical of filamentous fungi (Edelmann and Staben, 1994) The intron structure for two *lol* genes, *lolC* and *lolD* is explored in this chapter.

In addition, an in depth gene family analysis was used to explore the origin of *lolC* and *lolD*. Based on BLAST comparisons, each of these secondary metabolism genes has significant similarity to a primary metabolism gene from a well studied gene family, *lolC* to an O-acetylhomoserine sulfhydrylase (*oahsh*) and *lolD* to an ornithine decarboxylase (*odc*). Interestingly, both of these gene families are pyridoxal-5-phosphate binding (PLP) enzymes. However, the two genes are not likely to be evolutionarily related as the PLP fold appears to have evolved independently at least five times, giving rise to five distinct superfamilies (Mehta and Christen, 2000; John, 1995; Denessiouk et al., 1999), all of which are present in every kingdom of life.

The following study provides extensive evidence to support the hypothesis that the *LOL* gene cluster is a novel trait which originated within fungal ancestors of the Neotyphodium and Epichloë endophytes, rather than being transferred in from another kingdom via HGT.

Materials and Methods

Preparation of fungal material, PCR and sequencing were performed as described in Chapter III.

Cloning of lolC and lolD primary metabolism paralogs

The primary metabolism paralog to *lolC* was amplified by degenerate PCR using primers designed from fungal O-acetylhomoserine sulfhydrylase (*oah*) genes. Amino acid sequences from *Neurospora crassa* XP_328091, *Magnaporthe grisea* EAA56840, *Emericella nidulans* P50125, and *Schizosaccharomyces pombe* NP_595189 were aligned by ClustalW (Thompson et al., 1994), and blocks were generated by GIBBS (Lawrence et al., 1993). Primers were then designed by CODEHOP (Rose et al., 1998) (http://searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) and the clamp was modified to match the closest *E. festucae* relatives. The resulting primers OAH Up 5' GAAGTCGGCGATGATATCATCDATRTGYTC and OAH Low 5' TCCCGTGCCGTCCCNATHTAYGCNAC were used under the following cycling conditions: 94°C denaturation for 30 s, 58°C annealing for 30 s, and 72°C extension for 2 1/2 min, repeated for 45 cycles, followed by a final 10 min extension. We amplified and cloned primary metabolism paralogs to *lolC* (putative O-acetylhomoserine sulfhydrylase; thus amino acid sequences referred throughout this chapter as pOAHsh) from *E. festucae* isolate 434 (EF_pOAHsh), *N. coenophialum* isolate e19 (NC_pOAHsh) and *Neotyphodium* sp. PauTG-1 isolate e55 (NP_pOAHsh).

The primary metabolism paralog to *lolD* was amplified in a similar manner. The specific amplification is described in detail in Chapter I (Kutil et al., 2004). Briefly, Odc amino acid sequences from *Tapesia yallundae*, *N. crassa*, *Paracoccidioides brasiliensis*, *Phaeosphaeria nodorum*, and *Coccidioides immitis* were aligned to design the primers ODC Up 5'-GCATCATCTATGCCCAGCCNTGYAARAC and ODC Low 5'-CGATGCCGTCGCAGGTNGGNCCCCA. A single 750 nucleotide product was amplified from *E. festucae* 189 genomic DNA using the conditions described above for *lolC* except that the amplification included 42 cycles of a 55°C annealing temperature and 1 min extension.

Compiling sequences from the Cys/Met Metab and Orn/DAP/Arg DC PLP-binding gene families

Gene family analyses include sequences compiled from the Pfam database version 15.0 (http://www.sanger.ac.uk/Software/Pfam/), SwissProt 45.3 /TrEMBL 28.3 (Boeckmann et al., 2003), Superfamily version 1.65 (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/), Genbank at NCBI (http://www.ncbi.nlm.nih.gov/) and the *Fusarium graminearum* Sequencing Project at the Broad Institute of MIT and Harvard (http://www.broad.mit.edu). We assembled 53 representative members of the Cysteine Methionine metabolism PLP-dependent enzyme (Cys/Met Metab PP; Pfam accession: PF01053) gene family for 31 species spanning the three superkingdoms (Fig. 11, Table 4). Similarily, we complied 49 representatives of the ornithine/diamino-

pimelate/arginine decarboxylase (Orn/DAP/Arg DC; pfam accession: PF00278) gene family from 33 species spanning the three superkingdoms (Fig. 14, Table 5).

Table 4
Species binomen and accession numbers for taxa in Cys/Met Metab PP gene family trees

Label	Full name	Accession Number [a]
AP_pCGS	*Aeropyrum pernix*	Q9YA71
AP_pCGL	*Aeropyrum pernix*	Q9YCN2
Atu_OAHsh	*Agrobacterium tumefaciens* str. C58	AAL42260
AT_CBL	*Arabidopsis thaliana*	P53780
AT_CGS	*Arabidopsis thaliana*	P55217
BH_CBL	*Bacillus halodurans*	Q9KCE3
BH_OAHsh	*Bacillus halodurans*	Q9K9P2
BH_CGL	*Bacillus halodurans*	Q9KC38
BH_CGS	*Bacillus halodurans*	Q9KCE4
BH_MGL	*Bacillus halodurans*	Q9KEQ0
BA_CBL	*Bordetella avium*	Q07703
CE_pCGL	*Caenorhabditis elegans*	P55216
EN_OAHsh	*Emericella nidulans*	P50125
EN_CBL	*Emericella nidulans*	AAB03241
EF_lolC	*Epichloë festucae*	EF012267
EF_pOAHsh	*Epichloë festucae*	EF015401
EC_CGS	*Escherichia coli*	P00935
EC_CBL2	*Escherichia coli*	P06721
FG_OAHsh	*Fusarium graminearum*	EAA67392
HI_CBL	*Haemophilus influenzae*	P44527
HI_CGS	*Haemophilus influenzae*	P44502
HP_CGS	*Helicobacter pylori*	P56069
HS_CGL	*Homo sapiens*	P32929
KL_MET17	*Kluyveromyces lactis*	Q92441
ML_CBL	*Mesorhizobium loti*	Q982C4
ML_OAHsh	*Mesorhizobium loti*	Q982W2
ML_OSHsh	*Mesorhizobium loti*	Q98BB2
ML_CGL	*Mesorhizobium loti*	Q98DX6
ML_MGL	*Mesorhizobium loti*	Q98K64
MM_CGL	*Mus musculus*	Q8VCN5
MLe_CGS	*Mycobacterium leprae*	P46807
MT_CGS	*Mycobacterium tuberculosis*	P66875

(Table 4 continued)

Label	Full name	Accession Number [a]
NC_lolC	*Neotyphodium coenophialum*	EF012268
NC_pOAHsh	*Neotyphodium coenophialum*	EF015402
NP_pOAHsh	*Neotyphodium* sp. PauTG-1	EF015403
NCr_pOAHsh	*Neurospora crassa*	EAA27038
NCr_CGL	*Neurospora crassa*	AAK94040
NCr_CBL	*Neurospora crassa*	AAK94039
PM_pOAHsh	*Prochlorococcus marinus*	CAE19101
PA_OSHsh	*Pseudomonas aeruginosa*	P55218
PP_MGL	*Pseudomonas putida*	P13254
RN_CGL	*Rattus norvegicus*	P18757
SC_CBS	*Saccharomyces cerevisiae*	P32582
SC_CGL	*Saccharomyces cerevisiae*	P31373
SC_pCBL2	*Saccharomyces cerevisiae*	P43623
SC_CBL1	*Saccharomyces cerevisiae*	P53101
SC_OAHsh	*Saccharomyces cerevisiae*	P06106
SC_YHR2	*Saccharomyces cerevisiae*	P38716
ST_CBL	*Salmonella typhimurium*	P18949
SP_OAHsh	*Schizosaccharomyces pombe*	O13326
SP_YFHE	*Schizosaccharomyces pombe*	O42851
SCr_CBL	*Streptococcus cremoris*	Q9RAS9
SCo_pCGL	*Streptomyces coelicolor*	Q59829
UM_hyp1	*Ustilago maydis*	EAK86447

[a] All accession numbers are from Genbank (accession numbers begin with at least two letters) or UniProt (accession numbers begin with P, Q, or O).

Table 5
Species binomen and accession numbers for taxa in Orn/DAP/Arg decarboxylase trees

Label	Species	Accession number[a]
CA_ODC	Candida albicans	P78599
CG_DCDA	Corynebacterium glutamicum	P09890
DC_ADC	Dianthus caryophyllus	Q96412
DM_ODC	Drosophila melanogaster	P40807
EN_ODC	Emericella nidulans	EAA59111
EF_ODC	Epichloe festucae	EF015404
EF_lolD	Epichloe festucae	EF012267

(Table 5 continued)

Label	Species	Accession number[a]
EC_ADC	Escherichia coli	P21170
EC_DCDA	Escherichia coli	P00861
EC_ALR1	Escherichia coli	Q8FB20
FG_ODC	Fusarium graminearum	EAA75548
GM_ADC	Glycine max (Soybean)	Q39827
HI_DCDA	Haemophilus influenzae	P19572
HP_DCDA	Helicobacter pylori	P56129
HS_ODC	Homo sapiens	P11926
HS_ODCAI	Homo sapiens	O14977
LD_ODC	Leishmania donovani	P27116
LE_ODC	Lycopersicon esculentum	O22616
LE_ADC	Lycopersicon esculentum	P49726
MJ_DCDA	Methanococcus jannaschii	Q58497
MM_ODCAI	Mus musculus	O35484
MM_ODC	Mus musculus	P00860
NC_lolD	Neotyphodium coenophialum	EF012268
NCr_ODC	Neurospora crassa	P27121
NS_ADC	Nicotiana sylvestris	O64453
NT_ODC	Nicotiana tabacum	P93351
PBCV_1	Paramecium bursaria chlorella vir1	Q84527
PSa_ADC	Pisum sativum (Garden pea)	Q43075
PF_ODC	Plasmodium falciparum	O15696
PA_DCDA	Pseudomonas aeruginosa	P19572
PS_TabA	Pseudomonas syringae	P31851
RN_ODCAI	Rattus norvegicus	Q63764
SC_ODC	Saccharomyces cerevisiae	P08432
SP_ODC	Schizosaccharomyces pombe	CAB45689
SCo_DCDA	Streptomyces coelicolor	Q9ZBH5
SY_DCDA	Synechocystis sp. (strain PCC 6803)	Q55484
SY_ADC	Synechocystis sp. (strain PCC 6803)	P74576
TB_ODC	Trypanosoma brucei	P07805
UM_ODC	Ustilago maydis	O14439
XL_ODC	Xenopus laevis	P27120

[a] All accession numbers are from Genbank (accession numbers begin with at least two letters) or UniProt (accession numbers begin with P, Q, or O).

Constructing maximum likelihood trees for PLP-binding gene families

Amino acid sequences from diverse members of each gene family were aligned using ClustalX with a gap opening penalty of 15 and a gap extension penalty of 0.3. The alignment was then submitted to the Rascal server (Thompson et al., 1994) for proofreading. After visual inspection, alignments were trimmed to remove regions at either end of the sequences that were not common to most species in the alignment. Maximum likelihood trees were generated using PROML (Felsenstein, 2005.) with a JTT model for amino acid replacement.

For the bootstrap values in Fig. 12 and Fig. 15, SeqBoot (Felsenstein, 2005.) was used to generate 500 replicated amino acid alignments then trees were generated in PROML. These bootstrap trees were combined to calculate a consensus tree using Consense (Felsenstein, 2005.).

Characterizing the relationship between lol genes and their closest PLP-binding paralogs

Bayesian trees with posterior probabilities were constructed using MrBayes 3.0 (Huelsenbeck and Ronquist, 2001) using specific settings for "mixed" amino acid model and the "gamma" protein set ("prset"). Two million cycles were conducted using a burn-in of 200,000 with a sample frequency of 100.

Intron boundaries for *lolC* and *lolD* were described previously (Spiering et al., 2005). To identify introns in the primary metabolism paralogs we relied upon published data or evaluated the sequences ourselves using FGENESH and classical GT-AG rules

for predicting fungal introns (Edelmann and Staben, 1994). Specifically, we predicted fungal introns which: 1) began with the consensus sequence GT(A/G)NGT allowing for an expected 10% C to T substitution in the sixth position; 2) end with a consensus (C/T)AG sequence (10% substitution to A in the first position, but the "AG" are 100% conserved in the second and third positons); 3) usually contain a consensus sequence of (A/G)CT(A/G)AC found at least 20 nucleotides from the beginning of the intron sequence; 4) fall within a general size range of 35-100 nucleotides but usually are between 50 to 65 nucleotides long. (Rep et al., 2006).

Results

Maximum likelihood and Bayesian trees of LolC and Cys/Met Metab PP gene family amino acid sequences

In protein homology searches the predicted LolC amino acid sequence is most similar to O-acetylhomoserine sulfhydrylase (OAHsh), a member of the cysteine methionine metabolism PLP-dependent (Cys/Met Metab PP) gene family. Fig. 11 shows a maximum likelihood tree representing the relationships among amino acid sequences from members of the Cys/Met Metab PP gene family from the eukaryote, bacteria, and archaea kingdoms. This gene family includes cystathionine gamma-lyase (CGL), cystathionine gamma-synthase (CGS), cystathionine beta-lyase (CBL), methionine gamma-lyase (MGL), O-succinylhomoserine sulfhydrylase (OSHsh), and O-acetyl-serine/homoserine sulfhydrylase (OAHsh).

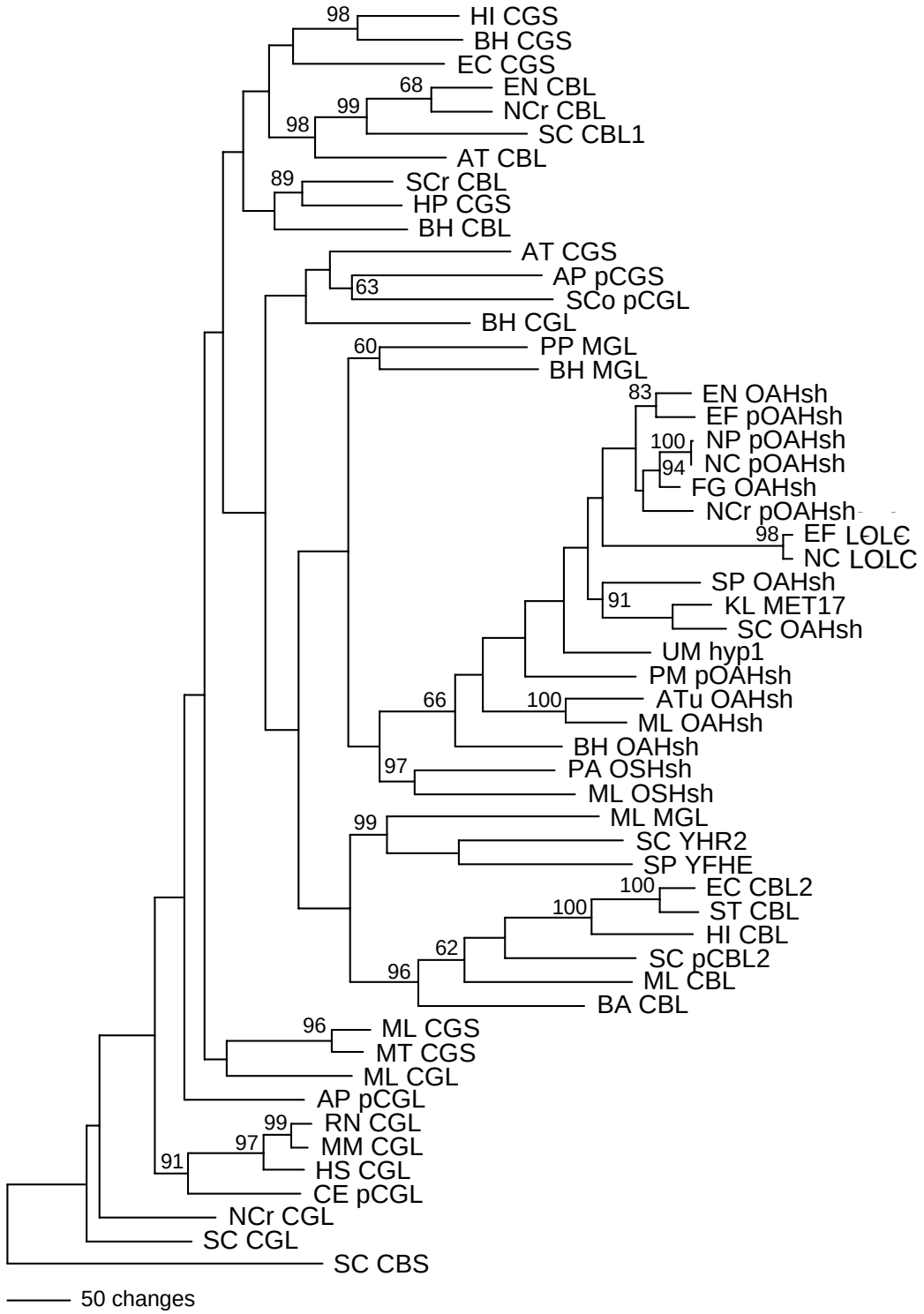

Fig. 11. Maximum Likelihood tree of Cys/Met Metab PP gene family sequences. Bootstrap values greater than 60 are indicated next to each node. Species abbreviations and protein accession numbers are provided in Table 4. Gene names are described in the results section.

Degenerate primers designed based on available fungal OAHsh sequences successfully amplified one primary metabolism paralog to *lolC* from *N. coenophialum* and *Neotyphodium* sp. PauTG-1 and a different primary metabolism paralog from *E. festucae* isolate 434. These two genes are only 70% identical at the nucleotide level but each is most similar to an OAHsh gene in BLAST searches. In the ML tree for the representatives of the Cys/Met Metab PP superfamily (Fig. 11) the LolC branch is a sister clade to filamentous fungal OAHsh amino acid sequences and yeast OAHsh proteins fall just outside those groups. Because the other members of the gene family are so diverged, further comparisons involved only the LolC sequences and OAHsh predicted amino acid sequences from fungi.

Fig. 12 shows a Bayesian tree of representative predicted fungal OAHsh and LolC amino acid sequences. The LolC amino acid sequences are more closely related to the OAHsh amplified from *E. festucae* isolate 434 and the OAHsh from *E. nidulans*. The OAHsh sequences amplified from *N. coenophialum* and *Neotyphodium* sp. PauTG-1 are more similar to the *F. graminearum* OAHsh.

Conservation of intron splice sites between lolC and fungal oahsh sequences.

Fig. 13 is a diagram of all intron positions found in *lolC* and its closest fungal *oahsh* paralogs, a total of 10 possible positions. All 5 *lolC* genes studied have 5 predicted introns located in identical positions (I1, I2, I4, I7, and I8). These five *lolC* introns are at the same positions as the five introns in the *E. nidulans oahsh*. The intron at position I8 is found only in *lolC* and the *E. nidulans oahsh*, while the other four intron

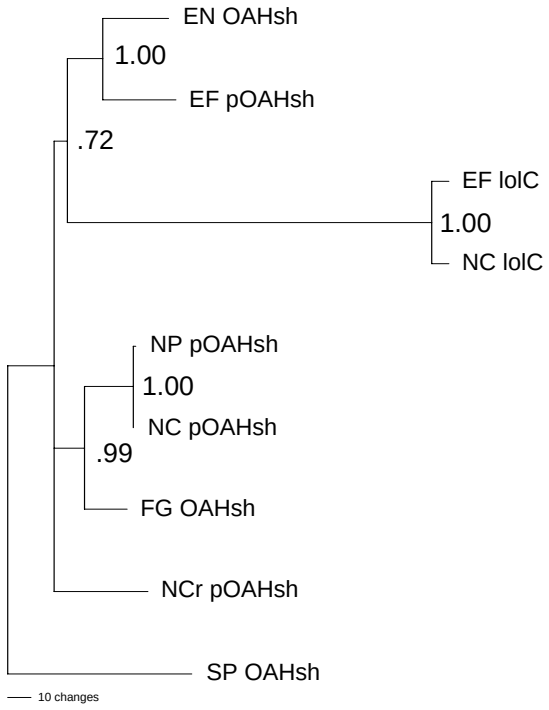

Fig. 12. Bayesian tree for endophyte LolC and fungal OAHsh gene family sequences. Posterior probabilities greater than 0.50 are indicated next to each node. Species abbreviations and protein accession numbers are provided in Table 4.

positons are found in *F. graminearum* and *N. crassa* as well. The I2 and I4 positions are also conserved in the predicted *oahsh* sequences amplified from the endophytes *N. coenophialum* and *Neotyphodium* sp. PauTG-1 but not the *E. festucae oahsh*. The *oahsh* from *N. coenophialum* and *Neotyphodium* sp. PauTG-1 also has one unique intron site (I9) and one intron site (I10) found in *F. graminearum* but none of the other model species analyzed. The *E. festucae oahsh* shares the I7 site with each of the species analyzed except for the other endophyte *oahsh* and has three unique intron sites (I3, I5, and I6). The conservation of intron positions suggests that *lolC* evolved from a gene similar to the *oahsh* from filamentous fungi.

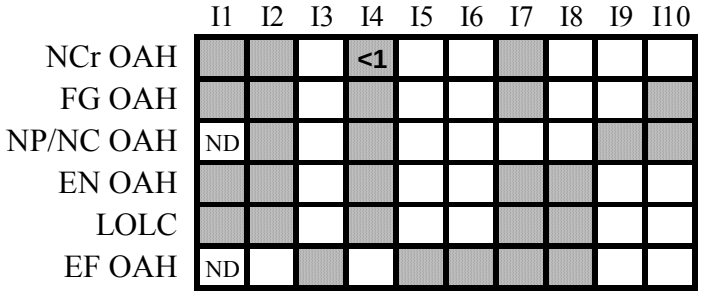

Fig. 13. Diagram of intron positions found in *lolC* and fungal *oahsh* sequences. Shaded boxes indicate that an intron is found at that location in a given species. "ND" indicates that we do not have sequence for that region, therefore the presence or absence of an intron cannot be determined. "<1" indicates that the NCr OAH sequence has what appears to be the I4 intron positioned one nucleotide upstream of the I4 intron in other sequences. Species abbreviations and protein accession numbers are provided in Table 4.

Maximim likelihood and Bayesian trees of LolD and Orn/DAP/Arg DC gene family amino acid sequences

Predicted LolD amino acid sequences are most closely related to eukaryotic ornithine decarboxylase (Odc) amino acid sequences. ODC is a member of the Orn/Dap/Arg decarboxylase family of PLP-dependent enzymes. Fig. 14 is the maximum likelihood gene tree comparing LolD predicted amino acid sequences to sequences for representative members of this gene family (Table 5). Ornithine decarboxylases (ODC) found in eukaryotes catalyze the conversion of ornithine into putrescine. Biosynthetic arginine decarboxylases (ADC) found in bacteria and plants catalyze the first step in putrescine synthesis from arginine, the transformation of arginine into agmatine. Prokaryotic diaminopimelate decarboxylases (DCDA) catalyze the final step of lysine biosynthesis, the transformation of diaminopimelic acid into lysine. The TabA protein

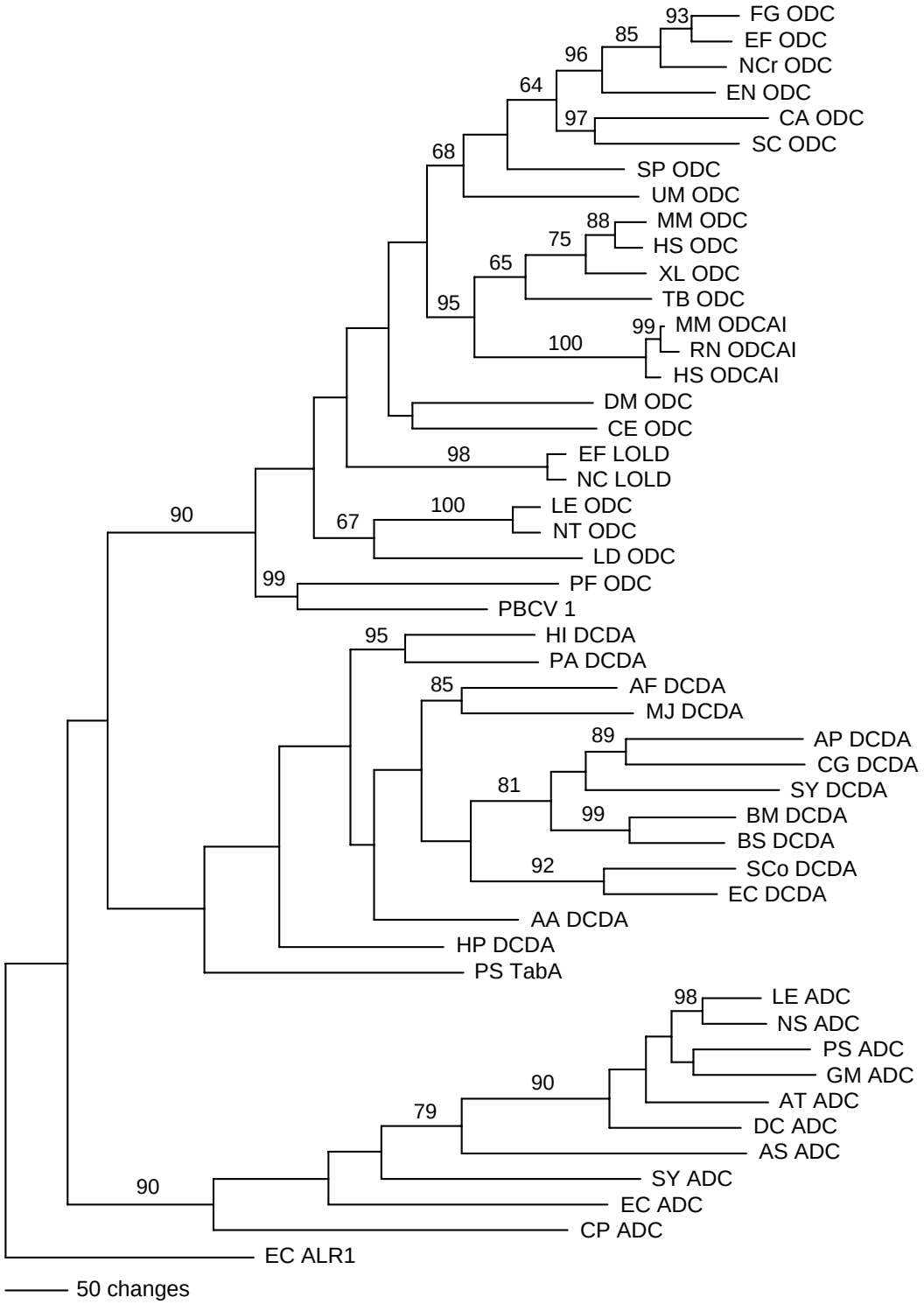

Fig. 14. Maximum Likelihood analysis of Orn/DAP/Arg gene family sequences. Bootstrap values over 60 are given on the branch leading to each node. Species abbreviations and protein accession numbers are provided in Table 5. Gene names are described in the results section.

from *Pseudomonas syringae* pv. tabaci is similar to DCDA and is hypothesized to be involved in the biosynthesis of tabtoxin. These proteins are all members of the group IV decarboxylases (Sandmeier et al., 1994), and though they likely share a common evolutionary origin, their sequence similarity is relatively low. LolD sequences are more similar to amino acid sequences encoding eukaryotic ODC enzymes than to the bacterial DCDA and the plant and bacterial ADC gene family members. Additionally it should be noted that bacterial ornithine decarboxylases are not represented here because they are not members of this family; rather, they are likely to represent an example of convergent evolution for the function (Mehta and Christen, 2000).

To further define the relationship of LolD with its closest paralogs, we constructed Bayesian tree including only eukaryotic Odc amino acid sequences with the protist *Plasmodium falciparum* as an outgroup (Fig. 15). In this tree LolD sequences group in a clade that is a sister clade to ascomycete Odc sequences, with basidiomycete sequences falling basal to this group.

Comparison of intron locations between LolD and fungal Odc sequences.

The close relationship of *lolD* to fungal *odc* genes is further supported by the presence of an intron at a site identical to all *odc* sequences analyzed from filamentous fungi. Not surprisingly, there is little sequence conservation within this intron between *lolD* and the *E. festucae odc* or across the different fungal species. All five *lolD* sequences also have two additional introns not found in any *odc* sequences analyzed.

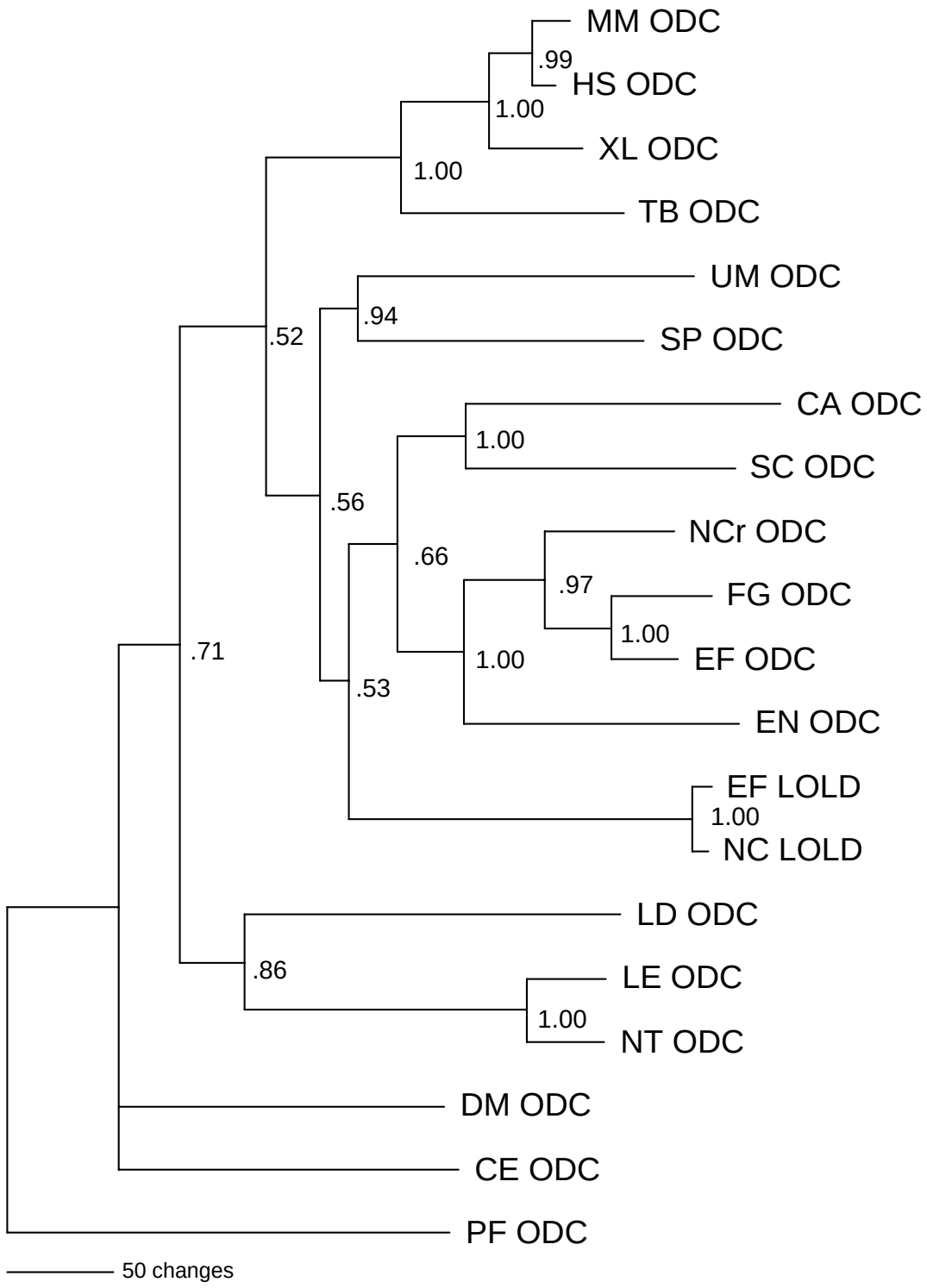

Fig. 15. Bayesian tree of eukaryotic ODC sequences. Posterior probabilities over 0.50 are given next to each node. Species abbreviations and protein accession numbers are provided in Table 5.

The positions of the three introns are the same in all five *lolD* sequences. While the sequence of each intron is highly conserved across *lolD* sequences, the sequences of introns at different positions have no identifiable similarity to each other or to the more ancient conserved introns.

Discussion

Origin of LOL

As a surrogate for identifying the origin of *LOL* we selected two *lol* genes for phylogenetic analysis. Both genes, *lolC* and *lolD* seem likely to be loline specific innovations, since the closest homologs found in Genbank or available fungal databases are PLP-dependent primary metabolism genes involved in amino acid biosynthesis. Given representation of all the subfamilies of the PLP-binding enzymes in the archea, eubacteria and eukaryota superkingdoms (Mehta and Christen, 2000) and given their involvement in so many fundamental primary metabolism pathways, it has been asserted that these are among the most ancient enzymes (Alexander et al., 1994; Cunchillos and Lecointre, 2003, 2005). Generally, it is expected that the five subfamilies evolved independently (convergently) to use PLP (a derivative of vitamin B6) as a cofactor. All of the PLP-binding enzyme types (representing five different patterns of protein folding, resulting in functional constraints on each class) evolved prior to divergence of the last common ancestor of the major kingdoms. Subsequently, there has been much more recent evolution of substrate binding specificities (Denessiouk et al., 1999; Eliot and

Kirsch, 2004). In many eukaryotes it is not uncommon to find tissue specific expression of paralogous PLP-binding enzymes involved in primary metabolism pathways (Eliot and Kirsch, 2004). Alternatively, some PLP-binding enzymes exhibit a remarkable amount of catalytic promiscuity, utilizing a single enzyme to catalyze different reactions (O'Brien and Herschlag, 1999). This sort of multi-functionality for a gene product may predispose the gene for subfunctionalization after a duplication event (Moore and Purugganan, 2005). There are numerous examples of genes encoding PLP-binding enzymes that have evolved roles in secondary metabolism in a variety of systems (Engst and Shaw, 1992; Koch et al., 2001). In some cases the same function evolved independently in different PLP fold classes (i.e. lineages). For example, the eukaryotic ornithine decarboxylase is evolutionarily related to the prokaryotic diaminopimelate decarboxylase (Martin et al., 1988) and is predicted to have evolved from a separate lineage than the bacterial ornithine decarboxylase (Mehta and Christen, 2000). Furthermore, there is at least one clear example of horizontal transfer of a PLP-binding enzyme between a eukaryote (an algae) and a virus (Shah et al., 2004).

The evolution of lolC from the Cys/Met Metab PP gene family

The largest PLP subfamily is particularly diverse, representing as much diversity within it as across any comparison of the five families of PLP-dependent protein folds (Mehta and Christen, 2000). In this alpha family (aka aminotransferase superfamily PF00155) there are examples of entire primary metabolism pathways that have evolved via gene duplication and recruitment of new substrate specificities (Cherest et al., 1993;

Gophna et al., 2005). The Cys/Met Metab PP family is part of this lineage of PLP-binding enzymes and includes cystathionine gamma-lyase, cystathionine gamma-synthase, cystathionine beta-lyase, methionine gamma-lyase, O-succinylhomoserine sulfhydrylase, and O-acetyl-serine/homoserine sulfhydrylase. Cystathionine gamma-lyase (CGL) catalyzes the final reaction in the transulfuration pathway leading from methionine to cysteine in eukaryotes, the transformation of cystathionine into cysteine, oxobutanoate and ammonia. Cystathionine gamma-synthase (CGS) catalyzes the first step in the biosynthesis of methionine from cysteine in bacteria, the conversion of cysteine and succinyl-homoserine into cystathionine and succinate. Cystathionine beta-lyase (CBL) catalyzes the second step in the biosynthesis of methionine from cysteine in bacteria, the conversion of cystathionine into homocysteine, pyruvate and ammonia. Methionine gamma-lyase (MGL) catalyzes the transformation of methionine into methanethiol, oxobutanoate and ammonia. (Barton et al., 1993; Ono et al., 1992) O-acetyl-homoserine sulfhydrylase (OAHsh) catalyzes the reaction of O-acetylhomoserine and homoserine into homocysteine. Additionally, it may participate in de novo methionine synthesis or an alternative homocysteine biosynthesis pathway in microorganisms that synthesize homocysteine primarily via cystathione (Yamagata, 1989; Sienko and Paszewski, 1998).

Our analysis indicates that *lolC* falls within this family; in fact, the *Neotyphodium uncinatum* sequences are included in this family in the Pfam database (Pfam accession numbers Q8JOB2_NEOUN and Q5MNH8_NEOUN). Fig. 12 demonstrates that *lolC* is most closely related to the O-acetylhomoserine sulfhydrolases

(*oahsh*) of filamentous fungi; *oahsh* sequences from some yeasts, more distantly related fungal ascomycetes, fall outside this clade. This is consistent with the hypothesis that *lolC* evolved from a duplicated *oahsh* in an ancestor of the endophytes.

However, further comparison of the LolC predicted amino acid sequences with only these OAHsh paralogs, reveals a more complicated pattern of relatedness among the predicted OAHsh genes amplified for these endophyte species. The putative OAHsh amplified from *E. festucae* is quite similar to the *E. nidulans* OAHsh, while the putative OAHsh from the two other endophyte species showed greater similarity to the *F. graminearum* OAHsh. Taken together with the intron splicing site data (Fig. 13), there is no simple explanation for these patterns of relatedness. BLAST searches of the available filamentous fungal genomes only yielded one *oahsh* per species. Thus, these different paralogs do not appear to co-occur in those fungal genomes. However, analysis of gene sequences from a single isolate from each species does not preclude the possibility that these two types of *oahsh* genes do exist within fungal species, perhaps maintained by balancing selection (Ward et al., 2002). Nonetheless, based on established phylogenies of these species, the tree reflects what we might have expected for the relationship among *oahsh* sequences with the only exception being the *oahsh* sequence we amplified from *Epichloë festucae* (434).

The *Epichloë festucae* genome is currently being sequenced. The isolate used for sequencing is a progeny of isolate 434. Perhaps, when the sequence of the genome is available it will shed new light on the variation of *oahsh* within the species or within the genome. We might expect to find the coexistence of two very different copies of the

gene, whereby the copy that amplified for our study is simply more similar to the primers we used, while the other copy would be more similar to the version in the other endophytes and model filamentous fungi closely related to the endophytes. Alternatively we might find only one copy, identical to that from 434, especially since the isolate is a progeny of 434. The least likely alternative is that we would find one copy that is a version more similar to that found in the other endophytes studied here, which would mean that the gene copy was from the other parental isolate in the cross with 434. Such an unexpected discovery would support the balancing selection hypothesis.

The evolution of lolD from the Orn/DAP/Arg DC gene family

Eukaryotic *odc* genes are members of one of the smallest lineages of PLP-binding proteins, the alanine racemace family (aka. amino acid decarboxylase group IV) (Mehta and Christen, 2000). The entire family consists of three members, ODC, DCDA, and ADC, all of which are included in our analysis (Fig. 14). As mentioned above genes encoding two of these enzymes, ADC and ODC have evolved independently in different PLP-binding enzyme lineages (eukaryotic versions are in this alanine racemace family, while the prokaryotic enzymes performing these functions evolved from a common ancestor in the alpha family; Mehta and Christen, 2000). Fig. 14 shows that the protein predicted from *lolD* is more closely related to the ODC gene family sequences than to the ADC or DCDA sequences. Once again, when we remove the more divergent sequences, the use of more informative characters allows for better resolution of LolD within the Odc lineages. In the Bayesian analysis in Fig. 15, LolD falls out with fungal

Odc as opposed to the mammalian or plant Odc. LolD even appears more closely related to the Odc sequences from filamentous fungi than does the Odc from the basidiomycete *Ustilago maydis*.

Intron analysis provides further support for the evolution of *lolD* via duplication and divergence of a fungal *odc*. Each of the filamentous fungal *odc* sequences analyzed contains only one intron, which is found in the same position in each of these sequences as well as *lolD*. This type of conservation of intron structure is usually indicative of shared ancestry (Long et al., 1995; de Souza et al., 1996a; de Souza et al., 1996b). While the topology of the *odc* gene family tree does not place *lolD* as precisely derived from endophyte ancestors as the *oahsh* tree shows for *lolC*, this could likely be explained by an increased divergence since the duplication event that gave rise to *lolD*. The divergence of *lolD* from *odc* is further explored in Chapter V. Based on these analyses, there is clear support for the evolution of both *lolC* and *lolD* as the result of the duplications of *oahsh* and *odc* genes, respectively, within fungal ancestors. Once again, this is consistent with the hypothesis that *LOL* evolved via recruitment of gene duplication and neofunctionalization events within the genome of a common ancestor to the endophytes.

CHAPTER V

FUNCTIONAL CHARACTERIZATION OF ENDOPHYTE-DERIVED LOLD AND ODC IN A MODEL FILAMENTOUS FUNGAL SYSTEM

Introduction

Neurospora crassa has been used as a model for filamentous fungi and eukaryotic genetics for over sixty years (Davis, 2000). In addition to its success as a classical genetic system it is becoming a model for genome analyses as well, particularly since the release of its genome sequence in 2001. Fortunately, among the many mutants previously characterized and publicly available are several strains with the ornithine decarboxylase gene, *spe*-1, knocked-out. According to previous studies in *N. crassa*, *spe-1* encodes an ornithine decarboxylase that catalyzes the initial rate-limiting step in polyamine biosynthesis, the conversion of ornithine to putrescine, which eventually leads to the formation of spermidine and spermine (Williams et al., 1992). The availability of such mutants provides a simple and elegant system to evaluate the functional overlap of the endophyted-derived ODC and LOLD (the protein product of the ornithine decarboxylase-like gene, *lolD*) with *N. crassa* ODC.

As described previously, *lolD* is a secondary metabolism gene paralogous to the primary metabolism gene *odc*. Unlike *lolC* which has multiple paralogs in each genome, each of the filamentous fungal genomes analyzed in this study appear to have only one

primary metabolism *odc*. Since the *odc* knock-out mutant is available for *N. crassa*, this allows for complementation studies of *lolD*.

Based on the proposed biochemical pathway for loline biosynthesis (Fig. 16; Spiering et al., 2005), L-proline and L-homoserine may be precursors that feed into the pathway for norloline, one of the loline alkaloids.

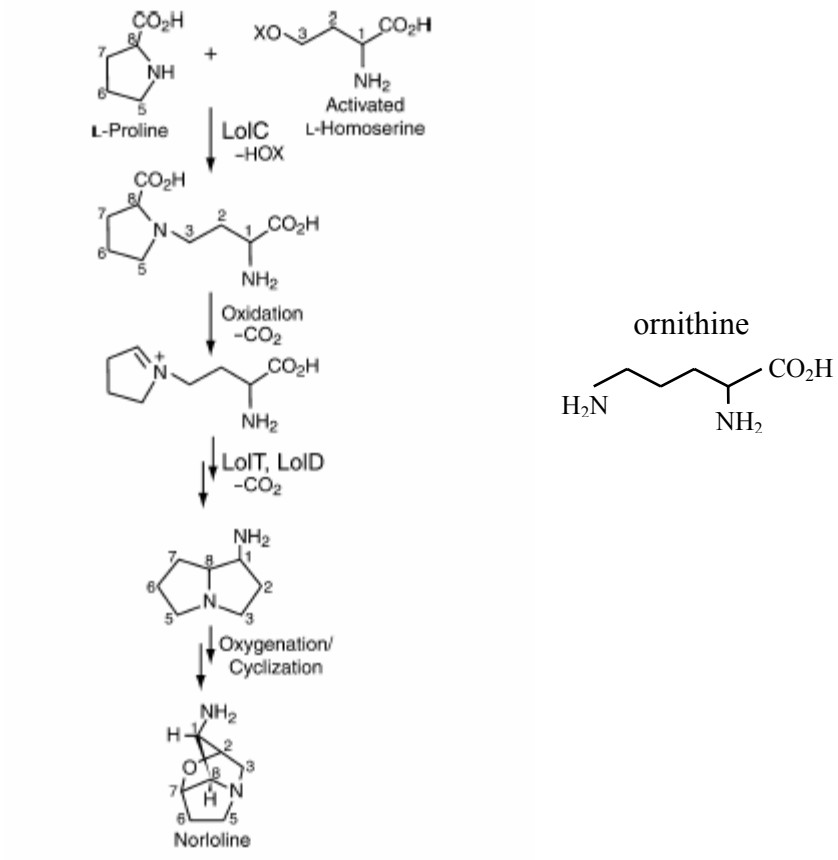

Fig. 16. Summary of a proposed biosynthetic pathway for one of the loline alkaloids. (Adapted from Spiering et al., 2005). Predicted roles of some gene products are indicated. On the right, the structure of ornithine is given next to the proposed substrate for LOLD.

A proposed substrate in the pathway which may be decarboxylated by LOLD appears very similar to ornithine, but has a ring structure on the end opposite the site of decarboxylation. To see whether this seems a logical prediction based on the evolutionary divergence of *lolD* from *odc*, we constructed a protein alignment and analyzed the conservation of important individual amino acid residues between LolD and several eukaryotic Odc sequences. Previous work has identified 26 amino acid residues implicated in the substrate binding domain, the pyridoxal phosphate binding domain and dimer stabilization in eukaryotic ODCs (Takatsuka et al., 2000). Our analysis here shows that, as expected, many of these residues are also conserved in LolD, but, perhaps more interestingly, a few are not. This further supports our hypothesis that *lolD* is the result of a duplication of *odc* which has neofunctionalized to perform a similar biochemical function on a different substrate.

The primary objective of the next experiment was to determine whether or not the *lolD* gene has diverged to the point that the LOLD protein it encodes is no longer able to complement the function of ODC. To this aim, I have ectopically integrated the complete coding region from *N. coenophialum lolD* under the control of a ribosomal protein (RP27) promoter (Lu et al., 2004) into *N. crassa* strains containing a loss of function *spe-1* mutation. Also, I used the same system to integrate a cassette containing the *E. festucae odc* and the *N. crassa odc* as positive controls. To evaluate the ability of LOLD to functionally complement ODC, the transformed strains are grown on media with and without the supplement spermidine (which restores growth in the *spe-1* mutant) as well as cadaverine, homoserine and proline to test for partial restoration of function.

Though not conclusive, these initial experiments demonstrate that only the strains transformed with endophyte *odc*, not *lolD* are functionally complementing the *N. crassa odc* mutant. In addition to serving as an appropriate control, this complementation by the endophyte *odc* is the first example of the successful expression of an *E. festucae* gene in *N. crassa*.

Materials and Methods

Preparation of fungal cultures

Neurospora crassa strain FGSC 4266 (*spe-1, inl*) was obtained from the fungal genetics stock center (Davis et al., 1987). This strain contains a *spe-1* allele 462JM which has a UV induced mutation in the ornithine decarboxylase gene (*odc*). This mutant requires that the media be supplemented with 50 μg/mL spermidine for growth (Eversole P., 1985). The strain also contains an inositol (*inl*) mutation which affects the myoinositol-1-phosphatase enzyme and requires that media be supplemented with inositol, 50 μg/mL, for proper growth in the media used in these experiments. (Davis et al., 1987)

Standard fungal media consisting of 1X Vogel's (V) and 2% sucrose (S) was supplemented with 50 μg/mL inositol (I), 50 μg/mL spermidine (Spe), cadaverine (Cad), 500 μg/mL proline (Pro), 200 μg/mL homoserine (Hse) and/or 200 μg/mL Hygromycin (Hyg; "Hyg 250" denotes 250 μg/mL Hygromycin B) as a selectable marker as

indicated. Chemicals and reagents were obtained from Sigma-Aldrich except for which purchased from VWR and supplied by EMD Biosciences Inc.

Preparation of N. crassa conidia for transformation

A sterile loop of *N. crassa* strain FGSC 4266 was inoculated into two 250 mL flasks containing 50 mL of V/S/I/Spe (Vogel's, sucrose, inositol, spermindine) with 1.5 % agar. Cultures were grown on the bench top for 2-3 weeks until bright orange conidia became abundant. Conidia were harvested from each flask using three washes of 15 mL 1M sorbitol filtered through sterile miracloth into 50 mL tubes. Tubes were centrifuged at 1500 rpm for 5 minutes. Liquid was discarded and each pellet was resuspended in 15 mL 1M sorbitol. Tubes were combined then centrifuged again and resuspended in 30 mL 1M sorbitol. After a total of 3 washes, approximately 200 µL of 1M sorbitol was added to the pellet, yielding a total volume of approximately 600 µL. An approximately 150 µL aliquot of conidia was used in each transformation.

Preparation and transformation of plasmid DNA into N. crassa

A 2 µg aliquot of each plasmid was digested with *Swa*I at 25ºC for 4 hrs. The digest was phenol/chloroform extracted, ethanol precipitated and resuspended in 20 µL water (Sambrook et al., 1989). A 10 µL aliquot was combined with 150 µL conidia into each cuvette and electroporated using 1500 volts, 25mA and 800 ohms. An additional 1 mL of 1M Sorbitol was added to the electroporated conidia and then they were stored on ice until spreading. Each transformation reaction was split evenly into two 25 mL

aliquots of top agar (with and without spermidine(spe)) and spread onto the respective agar plates (with and without spe). Plates were incubated at room temperature for 5-14 days before colonies were picked into 1mL of V/S/I/Spe/Hyg media in 10 by 75 mm culture tubes.

Homokaryon purification

Putative transformants were allowed to grow at room temperature for approximately one week before and during each stage of homokaryon purification (Davis and De Serres, 1970). Two rounds of homokaryon purification were completed as follows: 200 µL sterile water was added to a 10 mm slant of conidia; conidia were resuspended by vortexing; a sterile loop of wet conidia was added to 200 µL of water on a V/FGS/I/Spe/Hyg 250 plate and spread using a sterile glass rod; plates were incubated at room temperature for 4-8 days; individual colonies were picked into slants containing V/S/I/Spe/Hyg. After two rounds of purification, DNA was extracted as described below, and PCR amplification and sequencing were used to verify transformants.

Preparation of mycelia and DNA extraction

A loop of wet conidia was inoculated into 50 mL liquid Vogel's containing V/S/I/Spe/Hyg in a 250 mL beaker and shaken at 100rpm at 25ºC for 2-4 days (until a substantial mycelial mat was formed). The fungal mat was harvested and frozen at –20ºC for 3 hours then dried by lyophilizing overnight.

Under liquid nitrogen, the fungal mat was ground to a fine powder using a mortar and pestle. Next, the powder was resuspended in 2 mL of lysis buffer containing 50 mM Tris-HCl, 50 mM EDTA, 2%SDS and 20 µL B-mercaptoethanol. After incubating at 65ºC for 1 hr and centrifuging, the solution was extracted two times with phenol then once with cholorform, each time the aqueous (top) layer was harvested for the next step. After the final extraction the DNA was precipitated with 2 volumes of isopropanol and 200 µL of 0.3 M sodium acetate. The pellet was completely resuspended in 0.5 mL TE with 10 µg RnaseA and incubated at 37ºC for 1 hour. The phenol and chloroform extractions were repeated, two times with phenol and once with chloroform. The DNA was then precipitated from solution with 2 volumes of 95% ethanol and 1/10 volume of 0.3 M sodium acetate. The pellet was washed with 70% EtOH, allowed to air dry for 5 minutes, then resuspended in 100 µL TE and stored at 4ºC.

PCR amplification and sequencing were performed as described in Chapter III to verify transformants. The following primers were used for *N. crassa* transformants with *N. crassa odc*, *E. festucae odc* and *N. coenophialum lolD* coding regions, respectively: NCr odc Up AGGAACCCAATCTTCAAAATGGTTATGCCGACTGTTGTCTCCG, NCr odc Low AATGTTGAGTGGAATGATTTACAATCCCAAGAGCGCCATAGC, EF odc Up AGGAACCCAATCTTCAAAATGGCAACGGCTATCCTTG, EF odc Low AATGTTGAGTGGAATGATTTACATGTTCAAGAGGGCC, lolD Up AGGAACCCAATCTTCAAAATGACGACAGTCGTACGAG, and lolD Low AATGTTGAGTGGAATGATTTACCGTTCTCGGGGTGG.

Since these primers were also used to generate the transformation vectors, only the second half of each primer is designed to match the respective coding region. The first half of the primer was designed to allow incorporation into the transformation vector.

Experimental conditions

A total of 6 media types were tested with each putative transformant strain and the non-transformed controls. In all growth assays each strain was tested with a given medium in both a liquid culture and as a solid slant. Each transformant was inoculated into a V/S/I/Spe/Hyg slant, grown for approximately 7 days, resuspended by vortexing with 500 µL water, and transferred using a sterile cotton swab. After inoculation, experimental tubes were randomized in racks in a 25ºC diurnal incubator set for 12hr light, 12hr dark. After 7 days, the growth was recorded and pictures were taken as needed.

Protein sequence alignments

ODC gene family sequences were compiled as described in Chapter IV. Protein sequences were aligned using the Multalin web interface with a Blossum 62 symbol comparison table (Corpet, 1988). Alignments were exported as text files and manipulated in Microsoft Excel to generate figure shown.

Results

Sequences encoding the secondary metabolism *lolD* from *N. coenophialum* and the primary metabolism *odc* from *E. festucae* and *N. crassa* were each successfully incorporated into the *N. crassa* FGSC #4266 genome which contains a loss of function *spe-1* (*odc*) mutation. PCR and sequencing confirmed that these ectopic integrations include the coding region for the gene (start site to stop site including introns) under the control of a Magnaporthe ribosomal protein RP27 promoter (Lu et al., 2004). Four independent transformants were used in each experiment to evaluate the performance of each of the three genes (Table 6).

Table 6
Description of *N. crassa* transformants used for functional complementation experiments

Name	gene inserted
FGSC 4266[a]	none (odc-)
NBK 04	*N. crassa odc*
NBK 11	*N. crassa odc*
NBK 19	*N. crassa odc*
NBK 23	*N. crassa odc*
NBK 26	*E. festucae odc*
NBK 29	*E. festucae odc*
NBK 31	*E. festucae odc*
NBK 33	*E. festucae odc*
NBK 51	*N. coenophialum lolD*
NBK 67	*N. coenophialum lolD*
NBK 71	*N. coenophialum lolD*
NBK 74	*N. coenophialum lolD*

[a] *N. crassa* strain FGSC 4266 contains *spe-1* (ornithine decarboxylase) and *inl* (inositol) loss of function alleles.

Each of the strains in Table 6 was grown experimentally in liquid and solid tubes of Vogel's, Sucrose, inositol media supplemented with spermidine, cadaverine, proline, homoserine and hygromycin as indicated in Table 7.

Table 7
Growth of *lolD* and *odc* transformants on supplemented media. Growth is indicated by a +, while – indicates no growth on a given medium.

	N. crassa odc				*E. festucae odc*				*N. coenophialum lolD*				Controls (not transformed)	
	NBK 04	NBK 11	NBK 19	NBK 23	NBK 26	NBK 29	NBK 33	NBK 35	NBK 51	NBK 67	NBK 71	NBK 74	FGSC 4266A	FGSC 4266B
V/S/I	+	+	+	+	+	+	+	+	-	-	-	-	-	-
V/S/I/Cad	+	+	+	+	+	+	+	+	-[a]	-[a]	-[a]	-[a]	-[a]	-[a]
V/S/I/Pro	+	+	+	+	+	+	+	+	-	-	-	-	-	-
V/S/I/Hse	+	+	+	+	+	+	+	+	-	-	-	-	-	-
V/S/I/Spe	+	+	+	+	+	+	+	+	+	+	+	+	+[b]	+[b]
V/S/I/Spe/Hyg	+	+	+	+	+	+	+	+	+	+	+	+	-	-

[a] growth only on the surface of solid media with no aerial hyphae in solid or liquid cultures and almost no mycelial growth in liquid cultures.
[b] control strains showed relatively normal growth on the solid media supplemented with spermidine, but formed almost no arial hyphae in liquid media.

As expected, only strains complemented with an *odc* gene (from either *N. crassa* or *E. festucae*), not the *lolD* gene, were able to grow on media without spermidine. In addition to the PCR and sequencing, which verified transformants, the growth of all strains except the controls on media with hygromycin indicates that the ectopically integrated genes are being maintained in the genome of the transformants. Based on previous experiments characterizing the *spe-1* mutant allele in FGSC 4266, media supplemented with cadaverine is able to support a low level of growth in *spe-1* mutants without the addition of spermidine. Our results indicated that the *spe-1* mutants

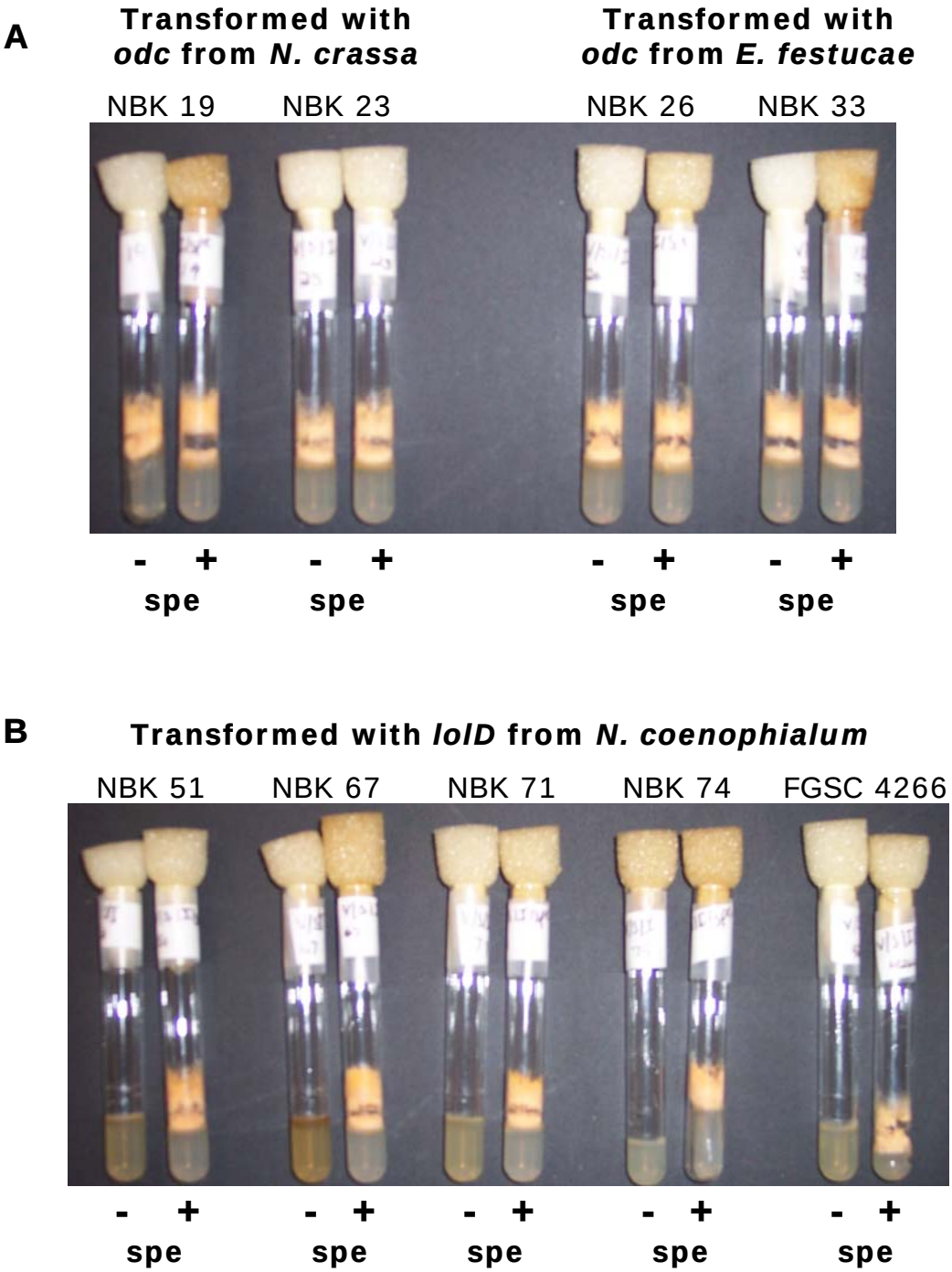

Fig. 17. Cultures of *N. crassa spe-1* (*odc*) mutant strains transformed with endophyte *odc* and *lolD* genes. Strains transformed with *Neurospora* or endophyte *odc* (A) or endophyte *lolD* (B) are grown on media with and without spermidine.

maintained this low level of growth on cadaverine supplemented media regardless of the presence or absence of *lolD*.

Fig. 17A further demonstrates that both the endophyte-derived *odc* and the *N. crassa odc* are able to restore the function of the *spe-1* mutation (i.e. growth on media not supplemented with spermidine). Fig. 17B demonstrates that the *N. crassa spe1-* strains transformed with *lolD* are not able to grow on media without spermidine. Furthermore, the darkening of the media in some tubes lacking spermidine is likely indicative of the yellow pigment secreted as part of the mutant phenotype (www.fgsc.net and personal communications from R. H. Davis and N. B. Raju). These yellow secretions are not seen in any of the strains that have the mutation complemented.

Based on the hypothesized biochemical pathway for the biosynthesis of lolines (Fig. 16), proline and homoserine are precursers for the loline alkaloid biosynthetic pathway in which LOLD plays a role (Spiering et al., 2005). However, in our analyses, none of the *N. crassa* strains transformed with *lolD* showed any partial complementation on media supplemented with either proline or homoserine. Fig. 16 also depicts a proposed substrate for LOLD. Comparison of this proposed substrate with the superimposed image of ornithine (the substrate for ODC) shows obvious similarities. The hypothetical substrate for LOLD appears to be an ornithine with a large ring structure attached to the end opposite the side which is decarboxylated. From this comparison, one could reasonably expect much of LOLD to be conserved with ODC.

Fig. 18. Alignment of partial Odc and LolD predicted amino acid sequences. Species abbreviations and accession numbers are described in Table 5. Conserved eukaryotic key residues identified by Takatsuka et al., 2000 are highlighted in yellow. Red highlights indicate the five residues in which the endophyte LolD sequences vary from the eukaryotic consensus.

Takatsuka et al. enumerates 26 amino acid residues implicated in the pyridoxal phosphate (PLP) binding domain, the substrate binding domain and in dimer stabilization in eukaryotic ODCs (Takatsuka et al., 2000). Analysis of predicted LolD amino acid sequences revealed they were identical for 21 of the 26 conserved amino acid residues (Fig. 18). Furthermore, the five residues that varied relative to eukaryotic Odc consensus were identical among the LolD sequences. Although these residues appear to

be spread throughout the sequence, when aligned using Cn3D to a three-dimensional computer image of a mouse ODC mmdb file (Kern et al., 1999), the five residues which are conserved in eukaryotic ODC but differ in LOLD actually fold into a very small proximity on one end of the substrate binding site (Fig. 19).

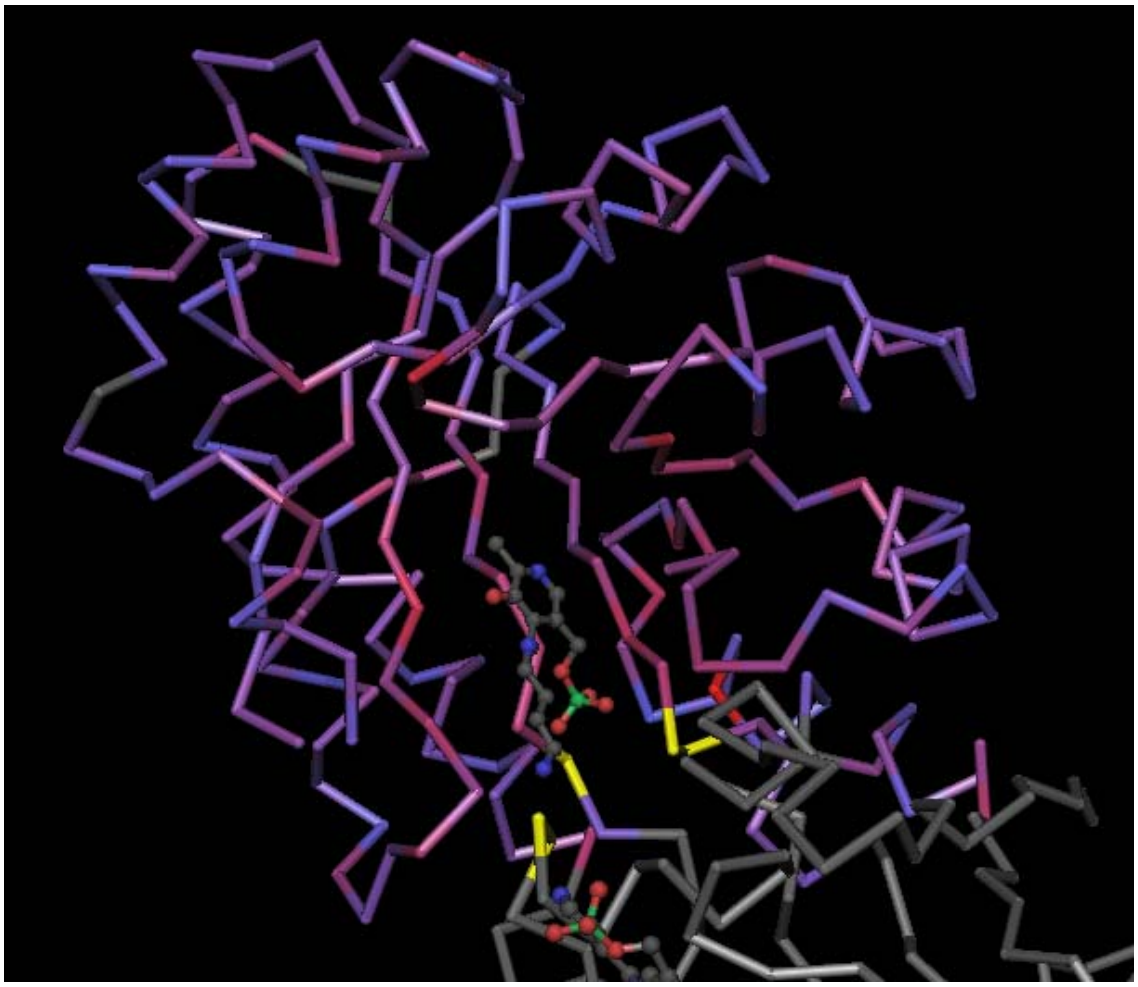

Fig. 19. Cn3D image of ODC. An ornithine ball and stick figure is shown in the predicted substrate binding site. Highlighted in yellow are the residues which are usually conserved across eukaryotic Odc but differ in LolD (those indicated in red in Fig. 18). Note that although there are five residues, only three protein segments are highlighted because two of the segments include more than one residue.

Discussion

Based on phylogenetic analysis, we have identified *lolD* as a secondary metabolism paralog of *odc* (Chapters III and IV). Previous work analyzing Odc has identified 26 key amino acid residues in eukaryotic ODC which are implicated in the pyridoxal phosphate (PLP) binding domain, the substrate binding domain and in dimer stabilization and are highly conserved among ODC in all eukaryotes (Takatsuka et al., 2000). When the LolD amino acid sequences are aligned with eukaryotic Odc sequences (Fig. 18) we see that 21 of these 26 residues are conserved in LolD as well. The 5 remaining residues differ from the Odc consensus but are conserved across LolD sequences. Furthermore, despite being separated by up to 80 residues in the predicted amino acid sequence, when aligned to the predicted three-dimensional ODC structure using Cn3D, these five residues appear to fold into very close proximity on one end of the substrate binding site (Fig. 19). Based on these data, our initial hypothesis predicts that LOLD has evolved a new function, which may be similar to that of ODC but acts on a different substrate, an intermediate in the biosynthesis of loline alkaloids (Fig. 16).

To evaluate this hypothesis experimentally we transformed a *N. crassa* mutant with a loss of function *odc* (*spe-1*) allele (FGSC 4266) with the *lolD* coding sequence from *N. coenophialum*. As positive controls, we also transformed the FGSC 4266 strain with *odc* from *E. festucae* and *N. crassa odc*. To evaluate the potential for LOLD and endophyte ODC to complement the function of *N. crassa* ODC, these transformants were grown on experimental media with various supplements as shown in Table 7. As expected only strains transformed with an *odc* gene were able to grow on media without

spermidine. These results clearly indicate that *lolD* is not complementing the loss of function of *odc*.

There are many possible explanations for negative results in the *lolD* complementation experiments. The constructs may lack the appropriate signals to ensure proper transcription, translation, processing or localization of the protein. Although independent replicate transformants were used in the experiments, each was generated from the same construct. Additional experiments would be required to verify the reason for the failure of the *lolD* transformants to functionally complement the *spe-1* mutation.

The success of the endophyte *odc* construct serves as an excellent positive control and proof of concept, as this is the first example of an *E. festucae* gene functionally complementing a mutation in *N. crassa*. This successful complementation of the *N. crassa odc* mutation by *E. festucae odc* is promising for the use of *Neurospora crassa* as a model system to further explore endophyte traits.

CHAPTER VI

CONCLUSION

Summary

LOL is a secondary metabolite gene cluster associated with the production of insecticidal loline alkaloids in grass endophytes (*Epichloë* and *Neotyphodium* spp.). The presence of a single copy of the gene cluster in the sexual species *E. festucae* provided an excellent genetic system to explore the evolution of this novel trait. A bacterial artificial chromosome (BAC) library was constructed for *Epichloë festucae*, allowing for map-based cloning of the *LOL* locus. I began by characterizing this BAC library. The 6,144 clone library with an average insert size of 87 kb represents at least 18-fold coverage of the 29 Mb genome. The original *LOL* polymorphism was mapped to the BAC 12-O9 in the *E. festucae* BAC library. Subcloning and sequencing of this BAC has identified two distinct regions. The first region, explored in Chapter II, spans at least 40kb and is present in both the Lol+ and the Lol- *E. festucae* parental isolates. This region is considered to be flanking the *LOL* locus and none of the genes in this region are implicated in the biosynthesis of loline alkaloids. In fact, this region contains many housekeeping genes which we expected to be maintained across diverse members of the ascomycota. Also in Chapter II, I evaluated the conservation of microsynteny between *E. festucae* and some model filamentous fungi by comparing sequence available from this region to publicly available fungal genome sequences. Orthologs to the 13

contiguous open reading frames (ORFs) identified in *E. festucae* are syntenic in *Neurospora crassa* and *Magnaporthe grisea* occurring in small sets of two, three or four colinear ORFs. This supports our prediction that many genes from *E. festucae* should exhibit a noticeable level of synteny when compared to the genome sequences of other model filamentous fungi.

I also used the BAC library to assemble a 110 kb contig spanning a putative ornithine decarboxylase gene (*odc*) (a primary metabolism paralog to *lolD*) and subsequently expanded it to 228 kb with a single walking step in each direction. None of the BACs which hybridized to the *oahsh* sequence from *E. festucae* appear to overlap with the *odc* contig. Therefore these two sequences do not appear to be within 200 kb of each other. However, it is possible that the *oahsh* version from *N. coenophialum* is found in a different location in the genome, perhaps much closer to *odc*. Or, the two genes may have at one time been much closer together, allowing for an initial duplication event to give rise to the genes which would evolve into *lol* genes.

The other end of BAC 12-O9, explored in Chapter III, spans a 50 kb region found only in the Lol+ parental isolate and progeny. Shotgun sequencing and assembly of this region revealed eight open reading frames implicated in the biosynthesis of loline alkaloids. Comparison with the two copies *LOL* identified in *N. uncinatum* allowed us to expand our comparisons to include 10 putative *lol* genes. In addition to the *E. festucae* sequence, I PCR amplified and sequenced the *LOL* locus from two additional asexual Lol+ endophytes, *Neotyphodium coenophialum* and *Neotyphodium* sp. PauTG-1. Comparative analysis including these five clusters from four different species is a

powerful tool for predicting both genetic architecture and the evolutionary history of *LOL*. In Chapter III, we use PhyloCon to compare predicted *lol* promoter regions, identifying four motifs conserved across the *lol* genes in all five clusters. Each motif has similarity to known fungal transcription factor binding sites in the TRANSFAC database. We speculate that the conservation of these motifs is further support for the hypothesis that *lol* genes are co-regulated.

Interestingly, the history of asexual *Neotyphodium* spp. includes multiple interspecific hybridization events. Comparing clusters from three *Neotyphodium* species and *E. festucae* allows us to determine which *Epichloë* ancestors are the most likely contributors of *LOL* in these asexual species. Since *E. festucae* is the only sexual species for which isolates able to produce loline alkaloids have been identified, it was originally expected to be a major contributor of *LOL* to the *Neotyphodium* lineages. However, some *Neotyphodium* species which are able to produce lolines show no evidence having *E. festucae* as an ancestor. This study provides evidence to support an alternative hypothesis that other *Epichloë* species may have, at one time, been able to produce loline alkaloids. For example, while no present day *E. typhina* isolates are known to produce lolines, our data suggest that the *E. typhina* ancestor(s) of three asexual endophyte species contained a *LOL* gene cluster which was passed on to *N. coenophialum*, *N. uncinatum*, and *Neotyphodium* sp. PauTG-1. Thus, Chapter III addresses a fundamental question in the evolution of fungal secondary metabolism by demonstrating that the polymorphism in production of lolines among endophyte species is likely due to the loss of the trait over time.

Chapter IV further explores the origin of the *LOL* gene cluster. Preliminary analysis of the introns in *lol* genes reveals a structure typically found in genes of filamentous fungi. Thus evolution from a ancestor within filamentous fungi seems far more likely than horizontal gene transfer from a prokaryote or more distantly related eukaryote. We completed an in depth gene family analysis to determine the most likely origin of two individual *lol* genes, *lolC* and *lolD*. The initial superfamily analyses revealed that each of these genes appears to have evolved as a secondary metabolism paralog to a primary metabolism gene, o-acetylhomoserine sulfhydrylase (*oahsh*) and ornithine dearboxylase (*odc*) respectively. These results support our predictions based on BLAST similarity searches. A more precise Bayesian analysis further indicated that each of the two *lol* genes most likely evolved from an ancestor within filamentous fungi. Therefore, the *LOL* gene cluster most likely originated in a fungal ancestor of endophytes, rather than being introduced via HGT from a more distantly related lineage.

The final study, Chapter V, analyzes the functional divergence of LOLD from its primary metabolism paralog ODC. For these experiments, we utilized an *N. crassa* strain available with a loss of function *spe-1 (odc)* allele. In independent experiments, I transformed this strain with a cassette containing either *lolD*, an endophyte derived *odc* or the *N. crassa odc*. As expected, both of the *odc* genes were able to restore normal growth, but transformation with the *lolD* gene did not restore growth.

In silica comparison of the predicted protein sequences provides further evidence for the neofunctionalization of *lolD* from *odc*. Prior analyses of eukaryotic Odc sequences predicted 26 amino acid residues important for substrate binding, dimer

stabilization, and the PLP fold which tended to be highly conserved across eukaryotic Odc. Alignment of the LolD predicted amino acid sequences with the eukaryotic Odc sequences reveals that 21 of these 26 residues are conserved across LolD as well. Interestingly, the five residues that differ from the eukaryotic consensus are conserved among the LolD sequences. Alignment to a three-dimensional ODC enzyme structure indicates that these five amino acid residues which differ in LOLD fold to one end of the substrate binding site. This supports our hypothesis that *lolD* has neofunctionalized to decarboxylate a substrate similar to ornithine except for a ring structure on the end of the substrate opposite the site of decarboxylation. Together with the gene family analyses in Chapter IV, this evidence indicates that *lolD* has evolved via duplication and neofunctionalization of a fungal *odc*.

Broader Context

The exploration of the *LOL* gene cluster is a significant contribution to the study of the evolution of fungal secondary metabolism gene clusters. To date, most evolutionary studies have focused on large multi-domain secondary metabolism genes such as PKS or NRPS, polyketide synthases and non-ribosomal peptide synthetases respectively. These genes are required components of many secondary metabolite gene clusters including clusters for the production of ergot alkaloids and sterigmatocystin (Brown et al., 1996; Coyle and Panaccione, 2005; Haarmann et al., 2005) and tend to exhibit rapid divergence, showing high sequence variability and domain shuffling even between closely related species. Furthermore, since most analyses of entire clusters

have utilized available genome sequences, the species used in the comparison tend to be distantly related. For example, the comparison of aflatoxin and sterigmatocystin gene clusters analyzes five *Aspergillus* spp. which diverged 25 million years ago (Ehrlich et al., 2005). The study that is perhaps most analogous to our exploration of *LOL* analyzes the trichothecene gene cluster from different chemotypes and closely related species of *Fusarium*. In that study, the phylogenies for the gene clusters from these very closely related isolates exhibited support for balancing selection acting to maintain the different chemotypes (Ward et al., 2002).

Finally, another study examined a cluster of four genes associated with lolitrem biosynthesis from both *E. festucae* and *N. lolii*, its asexual derivative. (Young et al., 2005). As expected, the genes were highly conserved, but the study also identified a very large retrotransposable element between two of the lolitrem cluster genes in *N. lolii*. As shown here for *LOL*, the retrotransposable element is highly degraded and no longer likely to be functional; however, unlike our results, the lolitrem gene cluster from the asexual species shows more evidence of corruption by retrotransposable elements than the sexual species, *E. festucae*. This is the opposite of what we observed for the *LOL* cluster. As more fungal genome sequences are released, the exploration of many fungal secondary metabolism gene clusters will become more feasible.

However, we must realize that understanding of evolutionary relationships may require analysis of even more closely related genomes than generally targeted in major sequencing initiatives.

Future Directions

We have come a long way in our understanding of the evolution of the *LOL* gene cluster. With the coming release of the *E. festucae* genome sequence, many of the questions raised in this study may find answers. Perhaps we will see evidence of a second primary metabolism *oahsh* gene or even a small region with a set of genes paralogous to *lol* genes which have been duplicated in a single event. Many of the individual *lol* genes appear to be members of large gene families with multiple members present throughout the genome, so mapping by other methods becomes exponentially difficult. The genome sequence will allow for a more expanded comparative analysis with other available genomes as well.

Further functional analysis of individual *lol* genes and their paralogs would be interesting as well. For example, the highly unusual presence of different *oahsh* genes in separate lineages of closely related endophytes raises the question whether or not these genes have subfunctionalized and are both being maintained in individual genomes or populations by some sort of balancing selection. Our success complementing the *N. crassa* mutant *odc* with endophyte *odc* suggests that *Neurospora* may serve as an excellent model system to further explore the function of endophyte genes; of course, RT-PCR, expression, and protein localization experiments should be included in more comprehensive experiments to allow for better interpretation of negative data.

As more genome sequences are released, it is possible, albeit unlikely, that we will find other clusters similar to *LOL*, especially in other closely related endophytes. Other members of the *Clavicipitaceae* (the same family as *Epichloë* and *Neotyphodium*)

are currently being sequenced including some that form insect-associations. It will certainly be interesting to see how the different host-associations affect the complement of secondary metabolites (loline alkaloids, ergot alkaloids, lolitrems and peramines among others) produced in the diverse lineages.

Other experiments are certain to shed more light on the evolution of the *LOL* cluster. In this study I present the regions flanking both sides of *LOL* in *E. festucae*. One obvious direction is the analysis of regions flanking *LOL* in other endophytes to ensure that *LOL* is in the same location for all endophytes. Perhaps we will find evidence of recent movement of *LOL* within the endophyte genome. Many interesting questions remain to be asked about the evolution of this novel trait.

REFERENCES

Al-Samarrai, T.H., and Schmid, J. 2000. A simple method for extraction of fungal genomic DNA. Lett. Appl. Microbiol. 30, 53-56.

Alexander, F.W., Sandmeier, E., Mehta, P.K., and Christen, P. 1994. Evolutionary relationships among pyridoxal-5'-phosphate-dependent enzymes. Regio-specific alpha, beta and gamma families. Eur. J. Biochem. 219, 953-960.

Arechavaleta, M., Bacon, C.W., Hoveland, C.S., and Radcliffe, D.E. 1989. Effect of the tall fescue endophyte on plant response to environmental stress. Agron. J. 81, 83-90.

Barton, A.B., Kaback, D.B., Clark, M.W., Keng, T., Ouellette, B.F., Storms, R.K., Zeng, B., Zhong, W., Fortin, N., Delaney, S., et al. 1993. Physical localization of yeast CYS3, a gene whose product resembles the rat gamma-cystathionase and *Escherichia coli* cystathionine gamma-synthase enzymes. Yeast 9, 363-369.

Belesky, D.P., and Malinowski, D.P. 2000. Abiotic stresses and morphological plasticity and chemical adaptations of *Neotyphodium*-infected tall fescue plants. In: Bacon, C.W., White, J.F. (Eds.), Microbial Endophytes. Marcell Dekker Inc., New York, pp. 455-484.

Belesky, D.P., Stringer, W.C., and Plattner, R.D. 1989. Influence of endophyte and water regime upon tall fescue accessions. II. Pyrrolizidine and ergopeptine alkaloids. Ann. Bot. 64, 343-349.

Berbee, M.L. 2001. The phylogeny of plant and animal pathogens in the Ascomycota. Physiol. and Mol. Plant Pathol. 59, 165-187.

Berbee, M.L., and Taylor, J.W. 2001. Fungal molecular evolution: gene trees and geologic time. In: McLaughlin, D.J., McLaughlin, E.G., Lemke, P.A. (Eds.), Systematics and Evolution. Springer-Verlag, Berlin, pp. 229-245.

Blankenship, J.D., Spiering, M.J., Wilkinson, H.H., Fannin, F.F., Bush, L.P., and Schardl, C.L. 2001. Production of loline alkaloids by the grass endophyte, *Neotyphodium uncinatum*, in defined media. Phytochemistry 58, 395-401.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S., and Schneider, M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365-370.

Bok, J.W., Balajee, S.A., Marr, K.A., Andes, D., Nielsen, K.F., Frisvad, J.C., and Keller, N.P. 2005. LaeA, a regulator of morphogenetic fungal virulence factors. Eukaryot. Cell 4, 1574-1582.

Bok, J.W., Hoffmeister, D., Maggio-Hall, L.A., Murillo, R., Glasner, J.D., and Keller, N.P. 2006. Genomic mining for *Aspergillus* natural products. Chem. Biol. 13, 31-37.

Brasier, C.M. 2000. The rise of the hybrid fungi. Nature 405, 134-135.

Brown, D. W., Yu, J. H., Kelkar, H. S., Fernandes, M., Nesbitt, T. C., Keller, N. P., Adams, T. H., and Leonard, T. J. 1996. Twenty-five coregulated transcripts define a sterigmatocstin gene cluster in *Aspergillus nidulans*. Proc. Natl. Acad. Sci. USA 93, 1418-1422.

Bush, L.P., Fannin, F.F., Siegel, M.R., Dahlman, D.L., and Burton, H.R. 1993. Chemistry, occurrence and biological effects of saturated pyrrolizidine alkaloids associated with endophyte-grass interactions. Agric. Ecosyst. Environ. 44, 81-102.

Bush, L.P., Wilkinson, H.H., and Schardl, C.L. 1997. Bioprotective alkaloids of grass-fungal endophyte symbioses. Plant Physiol. 114, 1-7.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol. 17, 540-552.

Challis, G.L., and Hopwood, D.A. 2003. Synergy and contingency as driving forces for the evolution of multiple secondary metabolite production by *Streptomyces* species. Proc. Natl. Acad. Sci. USA 100, 14555-14561.

Cheng, C., Kacherovsky, N., Dombek, K.M., Camier, S., Thukral, S.K., Rhim, E., and Young, E.T. 1994. Identification of potential target genes for Adr1p through characterization of essential nucleotides in UAS1. Mol. Cell Biol. 14, 3842-3852.

Cherest, H., Thomas, D., and Surdin-Kerjan, Y. 1993. Cysteine biosynthesis in *Saccharomyces cerevisiae* occurs through the transsulfuration pathway which has been built up by enzyme recruitment. J. Bacteriol. 175, 5366-5374.

Clark, L., and Carbon, J. 1976. A colony bank containing synthetic ColE1 hybrids representative of the entire *E. coli* genome. Cell 9, 91-99.

Clay, K. 1988. Clavicipitaceous fungal endophytes of grasses: co-evolution and change from parasitism to mutualism. In: Pirozynski, K.A., Hawksworth, D.L. (Eds.), Co-evolution of Fungi with Plants and Animals. Academic Press, London, pp. 79-105.

Clay, K. 1989. Clavicipitaceous endophytes of grasses: their potential as biocontrol agents. Mycol. Res. 92, 1-12.

Clay, K., and Cheplick, G.P. 1989. Effect of ergot alkaloids from fungal endophyte-infected grasses on fall armyworm (*Spodoptera frugiperda*). J. Chem. Ecol. 15, 169-182.

Clay, K., and Schardl, C.L. 2002. Evolutionary origins and ecological consequenes of endophyte symbiosis with grasses. Am. Nat. 160, 99-127.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16, 10881-10890.

Coyle, C. M., and Panaccione, D. G. 2005. An ergot alkaloid biosynthesis gene and clustered hypothetical genes from *Aspergillus fumigatus*. Appl. Environ. Microbiol. 71, 3112-3118.

Craven, K.D., Blankenship, J.D., Leuchtmann, A., Hignight, K., and Schardl, C.L. 2001a. Hybrid fungal endophytes symbiotic with the grass *Lolium pratense*. Sydowia 53, 44-73.

Craven, K.D., Hsiau, P.T.W., Leuchtmann, A., Hollin, W., and Schardl, C.L. 2001b. Multigene phylogeny of *Epichloe* species, fungal symbionts of grasses. Ann. Missouri Bot. Gard. 88, 14-34.

Cunchillos, C., and Lecointre, G. 2003. Evolution of amino acid metabolism inferred through cladistic analysis. J. Biol. Chem. 278, 47960-47970.

Cunchillos, C., and Lecointre, G. 2005. Integrating the universal metabolism into a phylogenetic analysis. Mol. Biol. Evol. 22, 1-11.

Dahlman, D.L., Siegel, M.R., and Bush, L.P. 1997. Insecticidal activity of N-formylloline. In: Proceedings of the XVIII International Grassland Congress, Winnipeg, Manitoba, Canada, pp. 13.15-13.16.

Davis, R.H. 2000. Neurospora contributions of a model organism. University Press, Oxford.

Davis, R.H., and De Serres, F. J. 1970. Genetic and microbial research techniques for *Neurospora crassa*. Methods Enzymol. 17A, 79-143.

Davis, R.H., Hynes, L.V., and Eversole-Cire, P. 1987. Nonsense mutations of the ornithine decarboxylase structural gene of *Neurospora crassa*. Mol. Cell Biol. 7, 1122-1128.

de Souza, S.J., Long, M., and Gilbert, W. 1996a. Introns and gene evolution. Genes Cells 1, 493-505.

de Souza, S.J., Long, M., Schoenbach, L., Roy, S.W., and Gilbert, W. 1996b. Intron positions correlate with module boundaries in ancient proteins. Proc. Natl. Acad. Sci. USA 93, 14632-14636.

Denessiouk, K.A., Denesyuk, A.I., Lehtonen, J.V., Korpela, T., and Johnson, M.S. 1999. Common structural elements in the architecture of the cofactor-binding domains in unrelated families of pyridoxal phosphate-dependent enzymes. Proteins 35, 250-261.

Diaz-Perez, S.V., Crouch, V.W., and Orbach, M.J. 1996. Construction and characterization of a *Magnaporthe grisea* bacterial artificial chromosome library. Fungal Genet. Biol. 20, 280-288.

Edelmann, S.E., and Staben, C. 1994. A statistical analysis of sequence features with genes from *Neurospora crassa*. Exp. Mycol. 18, 70-81.

Ehrlich, K.C., Yu, J., and Cotty, P.J. 2005. Aflatoxin biosynthesis gene clusters and flanking regions. J. Appl. Microbiol. 99. 518-527.

Eliot, A.C., and Kirsch, J.F. 2004. Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. Annu. Rev. Biochem. 73, 383-415.

Engst, K., and, and Shaw, P.D. 1992. Identification of a lysA-like gene required for tabtoxin biosynthesis and pathogenicity in *Pseudomonas syringae* pv. *tabaci* strain PTBR2.024. Mol. Plant Microbe Interact. 5, 322-329.

Eversole P, D.J., Menees T, and Davis RH. 1985. Structural gene for ornithine decarboxylase in *Neurospora crassa*. Mol. Cell Biol. 5, 1301-1306.

Farman, M.L., Tosa, Y., Nitta, N., and Leong, S.A. 1996. MAGGY, a retrotransposon in the genome of the rice bast fungus *Magnaporthe grisea*. Mol. Gen. Genet. 251, 665-674.

Faulkner, J.R., Hussaini, S.R., Blankenship, J.D., Pal, S., Branan, B.M., Grossman, R.B., and Schardl, C.L. 2006. On the sequence of bond formation in loline alkaloid biosynthesis. Chembiochem 7, 1078-1088.

Felsenstein, J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. (Distributed by the author. Department of Genome Sciences, University of Washington, Seattle).

Fidel, S., Doonan, J.H., and Morris, N.R. 1988. *Aspergillus nidulans* contains a single actin gene which has unique intron locations and encodes a g-actin. Gene 70, 283-293.

Fisher, F., and Goding, C.R. 1992. Single amino acid substitutions alter helix-loop-helix protein specificity for bases flanking the core CANNTG motif. EMBO J. 11, 4103-4109.

Fisher, R.A. 1930. The Genetical Theory of Natural Selection. Clarendon Press, Oxford.

Frugoli, J., and Harris, J. 2001. *Medicago truncatula* on the move! Plant Cell 13, 458-463.

Fu, Y.H., and Marzluf, G.A. 1990. nit-2, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein. Proc. Natl. Acad. Sci. USA 87, 5331-5335.

Funk, C.R., and White, J.R., Jr. 1997. Use of natural and transformed endophytes for turf improvement. In: Bacon, C.W., Hill, N.S. (Eds.), *Neotyphodium/*Grass Interactions. Plenum Press, New York, pp. 229-239.

Gophna, U., Bapteste, E., Doolittle, W.F., Biran, D., and Ron, E.Z. 2005. Evolutionary plasticity of methionine biosynthesis. Gene 355, 48-57.

Haarmann, T., Machado, C., Lübbe, Y., Correia, T., Schardl, C.L., Panaccione, D.G., and Tudzynski, P. 2005. The ergot alkaloid gene cluster in *Claviceps purpurea*: extension of the cluster sequence and intra species evolution. Phytochemistry 66, 1312–1320.

Hamer, L., Pan, H., Adachi, K., Orbach, M.J., Page, A., Ramamurthy, L., and Woessner, J.P. 2001. Regions of microsynteny in *Magnaporthe grisea* and *Neurospora crassa*. Fungal Genet. Biol. 33, 137-143.

Horowitz, N.H. 1945. On the evolution of biochemical syntheses. Proc. Natl. Acad. Sci. USA 31, 153-157.

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

Jacob, F., and Monod, J. 1961. Genetic regulatory mechanisms in the synthesis of proteins. J. Mol Biol. 3, 318-356.

John, R.A. 1995. Pyridoxal phosphate-dependent enzymes. Biochim. Biophys. Acta 1248, 81-96.

Johnson, M.C., Dahlman, K.L., Siegel, M.R., Bush, L.P., Latch, G.C.M., Potter, D.A., and Varney, D.R. 1985. Insect feeding deterrents in endophyte-infected tall fescue. Appl. Environ. Microbiol. 49, 568-571.

Judelson, H.S., and Randall, T.A. 1998. Families of repeated DNA in the oomycete *Phytophthora infestans* and their distribution within the genus. Genome 41, 605-615.

Kachroo, P., Chattoo, B.B., and Leong, S.A. 1994. Pot2, an inverted repeat transposon from *Magnaporthe grisea*. Mol. Gen. Genet. 245, 339-348.

Kachroo, P., Leong, S.A., and Chattoo, B.B. 1995. Mg-SINE: a short interspersed nuclear element from the rice blast fungus, *Magnaporthe grisea*. Proc. Natl. Acad. Sci. USA 92, 11125-11129.

Kang, S., Sweigard, J.A., and Valent, B. 1995. The PWL host specificity gene family in the rice blast fungus *Magnaporthe grisea*. Mol. Gen. Genet. 8, 939-948.

Keller, N.P., and Hohn, T.M. 1997. Metabolic pathway gene clusters in filamentous fungi. Fungal Genet. Biol. 21, 17-29.

Kern, A.D., Oliveira, M.A., Coffino, P., and Hackert, M.L. 1999. Structure of mammalian ornithine decarboxylase at 1.6 A resolution: stereochemical implications of PLP-dependent amino acid decarboxylases. Structure 7, 567-581.

Kimmons, C.A., Gwinn, K.D., and Bernard, E.C. 1990. Nematode reproduction on endophyte-infected and endophyte-free tall fescue. Plant Dis. 74, 757-761.

Klein, R.R., Morishige, D.T., Klein, P.E., Dong, J., and Mullet, J.E. 1998. High throughput BAC DNA isolation for physical map construction of sorghum (Sorghum bicolor). Plant Molec. Biol. Reporter 16, 351-364.

Koch, K.A., Capitani, G., Gruetter, M.G., and Kirsch, J.F. 2001. The human cDNA for a homologue of the plant enzyme 1-aminocyclopropane-1-carboxylate synthase encodes a protein lacking that activity. Gene 272, 75-84.

Kuldau, G.A., Tsai, H.F., and Schardl, C.L. 1999. Genome sizes of *Epichloe* species and anamorphic hybrids. Mycologia 91, 776-782.

Kutil, B.L., Liu, G., Verbalov, J., and Wilkinson, H.H. 2004. Contig assembly and microsynteny analysis using a bacterial artificial chromosome library for *Epichloë festucae*, a mutualistic fungal endophyte of grasses. Fungal Genet. Biol. 41, 23-32.

Lacey, J. 1991. Natural occurrence of mycotoxins in growing and conserved forage crops. In: Smith, J.E., Henderson, R.S. (Eds.), Mycotoxins and Animal Foods. CRC Press, Boca Raton, Florida, pp. 363-414.

Latch, G.C.M. 1993. Physiological interactions of endophytic fungi and their hosts: biotic stress tolerance imparted to grasses by endophytes. Agric. Ecosys. Environ. 44, 143-156.

Lawrence, C.E., Altschul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., and Wootton, J.C. 1993. Detecting subtle sequence signals: a Gibbs sampling strategy for multiple algnment. Science 62, 208-214.

Lawrence, J.G., and Roth, J.R. 1996. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143, 1843-1860.

Leuchtmann, A., Schardl, C.L., and Siegel, M.R. 1994. Sexual compatibility and taxonomy of a new species of *Epichloe* symbiotic with fine fescue grasses. Mycologia 86, 802-812.

Leuchtmann, A., Schmidt D, and Bush LP 2000. Different levels of protective alkaloids in grasses with stroma-forming and seed-transmitted *Epichloë/Neotyphodium* endophytes. J. Chem. Ecol. 26, 1025-1036.

Long, M., de Souza, S.J., and Gilbert, W. 1995. Evolution of the intron-exon structure of eukaryotic genes. Curr. Opin. Genet. Dev. 5, 774-778.

Lu, G., Filippi, C., Li, D., and Ebbole, D.J. 2004. Identification and characterization of secreted proteins in Magnaporthe grisea. In: Kawasaki, S. (Ed.), Rice Blast: Interaction with Rice and Control. Proceedings of the Third International Rice Blast Conference, Kluwer Academic Publishers, The Netherlands, pp. 47-55.

Lyons, P.C., Plattner, R.D., and Bacon, C.W. 1986. Occurrence of peptide and clavine ergot alkaloids in tall fescue grass. Science 232, 487-489.

Malinowski, D. 1995. Rhizomatous ecotypes and symbiosis with endophytes as new possibilities of improvement in competitive ability of meadow fescue (*Festuca pratensis*). Swiss Federal Institute of Technology, Zurich.

Malinowski, D., Leuchtmann, A., Schmidt, D., and Nösberger, J. 1997. Growth and water status in meadow fescue (*Festuca pratensis*) is affected by *Neotyphodium* and *Phialophora* endophytes. Agron. J. 89, 673-678.

Marks, S., Clay, K., and Cheplick, G.P. 1991. Effects of fungal endophytes on interspecific and intraspecific competition in the grasses *Festuca arundinacea* and *Lolium perenne*. J. Appl. Ecol. 28, 194-204.

Martin, C., Cami, B., Yeh, P., Stragier, P., Parsot, C., and Patte, J.C. 1988. *Pseudomonas aeruginosa* diaminopimelate decarboxylase: evolutionary relationship with other amino acid decarboxylases. Mol. Biol. Evol. 5, 549-559.

Mehta, P.K., and Christen, P. 2000. The molecular evolution of pyridoxal-5'-phosphate-dependent enzymes. Adv. Enzymol. Relat. Areas Mol. Biol. 74, 129-184.

Moon, C.D., Craven, K.D., Leuchtmann, A., Clement, S.L., and Schardl, C.L. 2004. Prevalence of interspecific hybrids amongst asexual fungal endophytes of grasses. Mol. Ecol. 13, 1455-1467.

Moore, R.C., and Purugganan, M.D. 2005. The evolutionary dynamics of plant duplicate genes. Curr. Opin. Plant Biol. 8, 122-128.

Nagylaki, T. 1984. Evolution of multigene families under interchromosomal gene conversion. Proc. Natl. Acad. Sci. USA 81, 3796-3800.

Nei, M., Gu, X., and Sitnikova, T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune†system. Proc. Natl. Acad. Sci. USA 94, 7799-7806.

Nurhayati, N., and Ober, D. 2005. Recruitment of alkaloid-specific homospermidine synthase (HSS) from ubiquitous deoxyhypusine synthase: does *Crotalaria* possess a functional HSS that still has DHS activity? Phytochemistry 66, 1346-1357.

O'Brien, P.J., and Herschlag, D. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. Chem. Biol. 6, R91-R105.

Ober, D. 2005. Seeing double: gene duplication and diversification in plant secondary metabolism. Trends Plant Sci. 10, 444-449.

O'Donnell, K.E., Kister, H.C., Cigelnik, E., and Ploetz, R.C. 1998. Multiple evolutionary origins of the fungus causing Panama disease of banana: concordant evidence from nuclear and mitochondrial gene geneologies. Proc. Natl. Acad. Sci. USA 95, 2044-2049

Ohta, T. 1990. How gene families evolve. Theor. Popul. Biol. 37, 213-219.

Ono, B., Tanaka, K., Naito, K., Heike, C., Shinoda, S., Yamamoto, S., Ohmori, S., Oshima, T., and Toh-e, A. 1992. Cloning and characterization of the CYS3 (CYI1) gene of *Saccharomyces cerevisiae*. J. Bacteriol. 174, 3339-3347.

Panaccione, D.G., Johnson, F.D., Wang, J., Young, C.A., Damrongkool, P., Scott, B., and Schardl, C.L. 2001. Elimination of ergovaline from a grass-*Neotyphodium* endophyte symbiosis by genetic modification of the endophyte. Proc. Natl. Acad. Sci. USA 98, 12820-12825.

Pedersen, C., Wu, B., and Giese, H. 2002. A *Blumeria graminis* f.sp. *hordei* BAC library – contig building and microsynteny studies. Curr. Genet. 42, 103-113.

Popay, A.J., and Rowan, D.D. 1994. Endophytic fungi as mediators of plant insect interactions. In: Bernays, E.A. (Ed.), Insect-Plant Interactions. CRC Press, Boca Raton, Florida, pp. 83-103.

Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. 2004. A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. Proc. Natl. Acad. Sci. USA 101, 8233-8238.

Raisbeck, M.F., Rottinghaus, G.E., and Kendall, J.D. 1991. Effects of naturally occurring mycotoxins on ruminants. In: Smith, J.E., Henderson, R.S. (Eds.), Mycotoxins and Animal Foods. CRC Press, Boca Raton, Florida, pp. 647-677.

Randall, T.A., and Judelson, H.S. 1999. Construction of a Bacterial Artificial Chromosome library of *Phytophthora infestans* and transformation of clones into *P. infestans*. Fungal Genet. Biol. 28, 160-170.

Read, J.C., and Camp, B.J. 1986. The effect of the fungal endophyte *Acremonium coenophialum* in tall fescue on animal performance, toxicity, and stand maintenance. Agron. J. 78, 848-850.

Rep, M., Duyvesteijn, R.G.E., Gale, L., Usgaard, T., Cornelissen, B.J.C., Ma, L., and Ward, T.J. 2006. The presence of GC-AG introns in *Neurospora crassa* and

other euascomycetes determined from analyses of complete genomes: implications for automated gene prediction. Genomics 87, 338-347.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Peitrokovski, S., McCallum, C.M., and Henikoff, S. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly-related sequences. Nucleic Acids Res. 26, 1628-1635.

Rosewich, U.L., and Kistler, H.C. 2000. Role of horizontal gene transfer in the evolution of fungi. Annu. Rev. Phytopathol. 38, 325-363.

Rowan, D.D., and Gaynor, D.L. 1986. Isolation of feeding deterrents against Argentine stem weevil from ryegrass infected with the endophyte *Acremonium loliae*. J. Chem. Ecol. 12, 647-658.

Rowan, D.D., and Latch, G.C.M. 1994. Utilization of endophyte-infected perennial ryegrasses for increased insect resistance. In: Bacon, C.W., White, Jr., J.F. (Eds.), Biotechnology of Endophytic Fungi of Grasses. CRC Press, Boca Raton, Florida,, pp. 169-183.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Plainview, New York.

Sandmeier, E., Hale, T., and Christen, P. 1994. Multiple evolutionary origin of pyridoxal-5'-phosphate-dependent amino acid decarboxylases. Eur. J. Biochem. 221, 997-1002.

Schardl, C.L. 2001. *Epichloë festucae* and related mutualistic symbionts of grasses. Fungal Genet. Biol. 33, 69-82.

Schardl, C.L., and Craven, K.D. 2003. Interspecific hybridization in plant-associated fungi and oomycetes: a review. Mol. Ecol. 12, 2861-2873.

Schardl, C.L., Leuchtmann, A., Chung, K.-R., Penny, D., and Siegel, M.R. 1997. Coevolution by common descent of fungal symbionts (*Epichloë* spp.) and grass hosts. Mol. Biol. Evol. 14, 133-143.

Schardl, C.L., Leuchtmann, A., Tsai, H.-F., Collette, M.A., Watt, D.M., and Scott, D.B. 1994. Origin of a fungal symbiont of perennial ryegrass by interspecific hybridization of a mutualist with the ryegrass choke pathogen *Epichloe typhina*. Genetics 136, 1307-1317.

Schuller, C., Brewster, J.L., Alexander, M.R., Gustin, M.C., and Ruis, H. 1994. The HOG pathway controls osmotic regulation of transcription via the stress response

element (STRE) of the *Saccharomyces cerevisiae* CTT1 gene. EMBO J. 13, 4382-4389.

Shah, R., Coleman, C.S., Mir, K., Baldwin, J., Van Etten, J.L., Grishin, N.V., Pegg, A.E., Stanley, B.A., and Phillips, M.A. 2004. *Paramecium bursaria* chlorella virus-1 encodes an unusual arginine decarboxylase that is a close homolog of eukaryotic ornithine decarboxylases. J. Biol. Chem. 279, 35760-35767.

Shizuya, H., Birren, B., Kim, U., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. 1992. Cloning and stable maintenance of 300-kilobase pair fragments of human DNA in *Escherichia coli* using an F-vector based vector. Proc. Natl. Acad. Sci. USA 89, 8794-8797.

Siegel, M.R., Latch, G.C.M., Bush, L.P., Fannin, F.F., Rowan, D.D., Tapper, B.A., Bacon, C.W., and Johnson, M.C. 1990. Fungal endophyte-infected grasses: alkaloid accumulation and aphid response. J. Chem. Ecol. 16, 3301-3315.

Sienko, M., and Paszewski, A. 1998. The metG gene of *Aspergillus nidulans* encoding cystathionine beta-lyase: cloning and analysis. Curr. Genet. 35, 638-46.

Soderlund, C., Humphray, S., Dunham, A., and French, L. 2000. Contigs built with fingerprints, markers, and FPC V4.7. Genome Res. 10, 1772-1787.

Spatafora, J.W., and Blackwell, M. 1993. Molecular systematics of unitunicate perithecial ascomycetes: the *Clavicipitales-Hypocreales* connection. Mycologia 85, 912-922.

Spiering, M.J., Moon, C.D., Wilkinson, H.H., and Schardl, C.L. 2005. Gene clusters for insecticidal loline alkaloids in the grass-endophytic fungus *Neotyphodium uncinatum*. Genetics 169, 1403-1414.

Spiering, M.J., Wilkinson, H.H., Blankenship, J.D., and Schardl, C.L. 2002. Expressed sequence tags and genes associated with loline alkaloid expression by the fungal endophyte *Neotyphodium uncinatum*. Fungal Genet. Biol. 36, 242-254.

Stahl, F.W., and Murray, N.E. 1966. The evolution of gene clusters and genetic circularity in microorganisms. Genetics 53, 569-576.

Stoltzfus, A., Logsdon, J.M., Jr., Palmer, J.D., and Doolittle, W.F. 1997. Intron "sliding" and the diversity of intron positions. Proc. Natl. Acad. Sci. USA 94, 10739-10744.

Sung, G.H., Spatafora, J.W., Zare, R., Hodge, K.T., and Gams, W. 2001. A revision of *Verticillium* sect. *Prostrata*. II. Phylogenetic analyses of SSU and LSU nuclear

rDNA sequences from anamorphs and teleomorphs of the *Clavicipitaceae*. Nova Hedwigia 72, 29-46.

Takatsuka, Y., Yamaguchi, Y., Ono, M., and Kamio, Y. 2000. Gene cloning and molecular characterization of lysine decarboxylase from *Selenomonas ruminantium* delineate its evolutionary relationship to ornithine decarboxylases from eukaryotes. J. Bacteriol. 182, 6732-6741.

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 weight matrix choice. Nucleic Acids Res. 22, 4673-4680.

Torres-Guzman, J.C., Xoconstle-Cazares, B., Guevara-Olvera, L., Ortiz, L., San-Blas, G., Dominguez, A., and Ruiz-Herrera, J. 1996. Comparison of fungal ornithine decarboxylases. Curr. Microbiol. 33, 390-392.

Tsai, H., Liu, J., Staben, C., Christensen, M.J., Latch, G.C.M., Siegel, M.R., and Schardl, C.L. 1994. Evolutionary diversification of fungal endophytes of tall fescue grass by hybridization with *Epichloe species*. Proc. Natl. Acad. Sci. USA 91, 2542-2546.

Tudzynski, B. 2005. Gibberellin biosynthesis in fungi: genes, enzymes, evolution, and impact on biotechnology. Appl. Microbiol. Biotechnol. 66, 597-611.

Tunlid, A., and Talbot, N.J. 2002. Genomics of parasitic and symbiotic fungi. Current Opin. Microbiol. 5, 513-519.

Walton, J.D. 2000. Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: an hypothesis. Fungal Genet. Biol. 30, 167-171.

Wang, T., and Stormo, G.D. 2003. Combining phylogenetic data with co-regulated genes to identify regulator motifs. Bioinformatics 19, 2369-2380.

Ward, T.J., Bielawski, J.P., Kistler, H.C., Sullivan, E., and O'Donnell, K. 2002. Ancestral polymorphism and adaptive evolution in the trichothecene mycotoxin gene cluster of phytopathogenic *Fusarium*. Proc. Natl. Acad. Sci. USA 99, 9278-9283.

West, C.P. 1994. Physiology and drought tolerance of endophyte-infected grasses. In: Bacon, C.W., White, Jr., J.F. (Eds.), Biotechnology of Endophytic Fungi of Grasses. CRC Press, Boca Raton, Florida, pp. 87-99.

Wilkinson, H.H., and Schardl, C.L. 1997. The evolution of mutualism in grass-endophyte associations. In: Bacon, C.W., Hill, N.S. (Eds.), *Neotyphodium/*Grass Interactions. Plenum Press, New York, pp. 13-25.

Wilkinson, H.H., Siegel, M.R., Blankenship, J.D., Mallory, A.C., Bush, L.P., and Schardl, C.L. 2000. Contributions of fungal loline production to protection from aphids in a grass-fungus mutualism. Molecular Plant-Microbe Interactions 13, 1027-1033.

Williams, L.J., Barnett, G.R., Ristow, J.L., Pitkin, J., Perriere, M., and Davis, R.H. 1992. Ornithine decarboxylase gene of *Neurospora crassa*: isolation, sequence, and polyamine-mediated regulation of its mRNA. Mol. Cell Biol. 12, 347-359.

Wingender, E., Chen, X., Fricke, E., Geffers, R., Hehl, R., Liebich, I., Krull, M., Matys, V., Michael, H., Ohnhauser, R., et al. 2001. The TRANSFAC system on gene expression regulation. Nucleic Acids Res. 29, 281-283.

Yamagata, S. 1989. Roles of O-acetyl-L-homoserine sulfhydrylases in microorganisms. Biochimie 71, 1125-1143.

Yates, S.G., Fenster, J.C., and Bartelt, R.J. 1989. Assay of tall fescue seed extracts, fractions, and alkaloids using the large milkweed bug. Journal of Agricultural and Food Chemistry 37, 354-357.

Young, C., Bryant, M., Christensen, M., Tapper, B., Bryan, G., and Scott, B. 2005. Molecular cloning and genetic analysis of a symbiosis-expressed gene cluster for lolitrem biosynthesis from a mutualistic endophyte of perennial ryegrass. Molecular Genetics and Genomics 274, 13-29.

Yu, J.H., and Keller, N. 2005. Regulation of secondary metabolism in filamentous fungi. Annu Rev Phytopathol 43, 437-458.

Zhang, H.-B., and Wing, R.A. 1997. Physical mapping of the rice genome with BACs. Plant Molecular Biology 35, 115-127.

Zhang, Y., Wilkinson, H.H., Keller, N.P., and Tsitsigiannis, D. 2004. Secondary metabolite gene clusters. In: An, Z. (Ed.), Handbook of Industrial Mycology. Marcel Dekker, Inc., New York, pp. 355-387.

Zhu, H., Choi, S., Johnston, A.K., Wing, R.A., and Dean, R.A. 1997. A large-insert (130 kbp) bacterial artificial chromosome library of the rice blast fungus *Magnaporthe grisea*: genome analysis, contig assembly, and gene cloning. Fungal Genet. Biol. 21, 337-347.

VITA

Brandi Lynn Kutil

253 Faculty Center
Texas A&M University Corpus Christi
Corpus Christi, TX 78412

B. S., Genetics, Texas A&M University, 1999
Ph.D., Genetics, Texas A&M University, 2006

Publications

Kutil, B. L., G. Liu, J. Verbalov, and H. H. Wilkinson. 2004. Contig assembly and microsynteny analysis using a bacterial artificial chromosome library for Epichloë festucae, a mutualistic fungal endophyte of grasses. Fung. Genet. Biol. 41:23-32.

Kutil, B. L., K. Y. Seong, and R. Aramayo. 2003. Unpaired genes do not silence their paired neighbors. Curr. Genet. 43:425-432.

Kutil, B. L., and C. G. Williams. 2001. Triplet-repeat microsatellites shared among hard and soft pines. J. Hered. 92:327-332.