GENOMICS-AIDED INSIGHTS INTO THE *EPICHLOË*-TURFGRASS SYMBIOSIS

by

KAREN V. AMBROSE MICHAEL

A dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

Written under the direction of

Prof. Faith C. Belanger

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

OCTOBER, 2014

ABSTRACT OF THE DISSERTATION

GENOMICS-AIDED INSIGHTS INTO THE *EPICHLOË*-TURFGRASS SYMBIOSIS

By KAREN V. AMBROSE MICHAEL

Dissertation Director:

Prof. Faith C. Belanger

Nearly all plants engage in some form of symbiosis with fungi. An association that is of broad significance in turf is the symbiosis between endophytic fungi of the genus *Epichloë* and cool season grasses. There has been longstanding interest in the relationship between turfgrass and their *Epichloë* endosymbionts as endophyte infection of turfgrass can result in increased plant vigor, and confer tolerance to biotic and abiotic stress. Thus the *Epichloë*-turfgrass interaction is a promising model for studies aimed at understanding beneficial symbiotic associations in general.

Fine fescues are valued as low maintenance turf spp. The *E. festucae* infection of strong creeping red fescue (*Festuca rubra*) is important because the symbiosis protects the host plant against fungal phytopathogens. The phytopathogens cause the devastating dollar spot, and red thread diseases. Remarkably, no other *Epichloë*-turfgrass interaction has

been reported to confer fungal disease resistance. The *E. festucae*-infected fescues also have insect resistance.

No study to date has investigated the *E. festucae-F. rubra* mutualism. The goal of this thesis was to use SOLiD-SAGE transcriptome analysis to identify candidate genes that may be important in the endophyte-grass symbiosis.

The results of a high-throughput quantitative differential gene expression study between plants infected with *E. festucae* vis-à-vis endophyte-free plants, in addition to the discovery and functional characterization of a horizontally transferred gene into *Epichloë* spp., and the examination of *E. festucae* salicylate hydroxylase gene are detailed in this thesis. The findings from this dissertation are expected to contribute to the knowledge on fungal endosymbiont-plant mutualism.

ACKNOWLEDGEMENT

First and foremost, I would like to express my utmost gratitude and appreciation to my advisor, Dr. Faith Belanger, for her unparalleled mentorship, guidance, support and patience. I would like to thank my committee members, Dr. Barbara Zilinskas, Dr. James White and Dr. Debashish Bhattacharya for their encouragement and time. A very special thanks to my colleagues in the Belanger research lab and Department of Plant Biology & Pathology for their friendship. I am very grateful to the Rutgers Center for Turfgrass Science for their generous financial support. I am immensely thankful for the endless support, encouragement and care from my family. Above all, I thank the Lord for every blessing, and the wonderful opportunities I have had at Rutgers during my time here.

"Nothing in life is to be feared. It is only to be understood."
- Marie Curie

TABLE OF CONTENTS

	Page
Abstract of the Dissertation	ii
Acknowledgment	iv
List of Tables	ix
List of Illustrations	x
Chapter 1	1
Introduction	1
Plant-fungi symbiosis	2
Significance of turfgrass	3
The strong creeping red fescue	4
Turfgrass-endophyte interaction	4
Genome sequence of *Epichloë*	7
Gene expression studies	8
Goals of this thesis	10
References	12
Chapter 2	17
SOLiD-SAGE of endophyte-infected red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins	17
Introduction	18
Materials and methods	20
Plant and fungal materials	20
RNA and DNA isolation	21
454/GS FLX Titanium sequencing	21
De novo 454 transcriptome assembly	22
SOLiD-SAGE sequencing	23

TABLE OF CONTENTS (cont.)

Page

Computational analysis of SOLiD-SAGE tags 24

GO annotation 25

RT-PCR 26

DNA sequencing of *Ef-AFP* splice variant 26

Phylogenetic analysis 27

Results 28

454 transcriptome sequencing 28

Differentially expressed plant host genes identified by SOLiD-SAGE sequencing 29

The most abundant host plant genes 31

Physiological categories of differentially expressed plant genes 32

E. festucae gene expression *in planta* 35

Endophyte transcripts putatively involved in plant defense evasion 38

Nutrient acquisition mechanisms *in planta* 39

Reactive oxygen species in the fungal-plant symbiosis 42

Absence of highly expressed transcripts for fungal alkaloids 42

The highly expressed *E. festucae* antifungal protein gene is not present in many *Epichloë* spp. 43

Expression of an alternatively spliced variant of the *E. festucae* antifungal protein gene 49

Antisense SAGE tags 51

Discussion 53

References 57

TABLE OF CONTENTS (cont.)

Page

Chapter 3 66

Horizontal gene transfer of a bacterial insect toxin gene into the *Epichloë* fungal symbionts of grasses 66

Introduction 67

Materials and methods 70

Plant and fungal materials 70

RNA and DNA isolation 70

Amplification and sequencing of *Epichloë mcf-like* genes and cDNAs 70

Expression of the *E. typhina* subsp. *poae* Ps1-*mcf* gene in *E. coli* 71

Insect assays to determine toxicity of *E. typhina* subsp. *poae* Ps1-Mcf protein 73

Schematic representation of *mcf* gene structure 74

Phylogenetic analysis 74

Results 75

Detection of a *makes caterpillars floppy-like* gene in *Epichloë* genomes 75

Estimation of the age of the *mcf-like* gene HGT event into *Epichloë* 79

The *Epichloë mcf-like* gene donor is likely a bacterium related to *Xenorhabdus* or *Photorhabdus* 87

Expression and activity of the *Epichloë mcf-like* gene 90

Discussion 95

References 98

Chapter 4 104

Functional characterization of salicylate hydroxylase from *Epichloë festucae* Rose City, a fungal endosymbiont of red fescue 104

Introduction 105

TABLE OF CONTENTS (cont.)

Page

Materials and methods 107

Plant and fungal materials 107

RNA and DNA isolation 107

Amplification and sequencing of *E. festucae* RC salicylate hydroxylase gene and cDNA 108

Expression of the *E. festucae* Rose City salicylate hydroxylase gene in *Escherichia coli* 109

Purification of the *E. festucae* Rose City salicylate hydroxylase protein in *E. coli* 110

Enzyme activity assays of *E. festucae* Rose City salicylate hydroxylase 112

Growth assays of *E. festucae* Rose City fungal endophyte on salicylate minimal medium 112

Nucleic and amino acid sequence alignments 113

Schematic representation of salicylate hydroxylase gene structure 113

Results 113

PCR of *E. festucae* Rose City salicylate hydroxylase coding sequence 113

Expression and purification of *E. festucae* Rose City salicylate hydroxylase protein 122

Detection of a persistent *Escherichia coli* metal-binding contaminating protein 125

Catalytic activity of *E. festucae* Rose City salicylate hydroxylase 127

Discussion 128

References 134

Chapter 5 141

Summary 141

LIST OF TABLES

Page

Table 2.1. Sequences of oligonucleotide primers used in this study 27

Table 2.2. Characteristics of the 454 sequences 29

Table 2.3. Characteristics of the SOLiD-SAGE tags 30

Table 2.4. Gene Ontology (GO) categorization of the 209 differentially expressed plant genes found by SOLiD-SAGE 34

Table 2.5. The 20 most abundant *E. festucae* transcripts found in the endophyte-infected plant S1139RC 36

Table 2.6. Accession numbers of the MCM7 and antifungal protein sequences used in the phylogenetic analyses presented in Figure 2.2 48

Table 3.1. Comparison of *Epichloë* Mcf-like amino acid sequences with those of the bacterial Mcf1, Mcf2, and FitD sequences 79

Table 3.2. Sources of DNA sequences used in this study 84

Table 3.3. Estimation of the divergence rate of the basal *Epichloë* spp. relative to *Atkinsonella hypoxylon* 85

Table 3.4. Estimation of the divergence times between the basal *Epichloë* spp. and *Claviceps purpurea* 86

Table 3.5. Estimation of the divergence time between *E. gansuensis* and *E. glyceriae* 87

Table 3.6. Cumulative numbers of black cutworm larvae deaths after each treatment 94

Table 4.1. NCBI accessions of ESTs containing *E. festucae* salicylate hydroxylase 116

Table 4.2. Mean activity of triplicate assays obtained using purified heterologously expressed *E. festucae* RC salicylate hydroxylase 127

LIST OF ILLUSTRATIONS

Page

Figure 2.1. Standard SOLiD-SAGE tag construction overview 24

Figure 2.2. The phylogenetic relationships of the MCM7 and antifungal protein coding sequences 47

Figure 2.3. Detection of an alternatively spliced variant of *Ef-AFP* 50

Figure 2.4. Gel analysis of *F. rubra* and *E. festucae* antisense transcripts 52

Figure 3.1. Gene structure of the *Epichloë mcf-like* genes and amino acid similarity of *E. typhina* subsp. *poae* Ps1-Mcf protein with the bacterial Mcf proteins 78

Figure 3.2. Rooted 50% majority rule maximum parsimony phylogenetic tree of the MCM7 DNA coding sequences 82

Figure 3.3. Rooted 50% majority rule maximum parsimony phylogenetic trees of the glutamate dehydrogenase, isoleucine tRNA synthase, and TSR1 DNA coding sequences 83

Figure 3.4. Rooted 50% majority rule maximum parsimony phylogenetic tree of bacterial Mcf proteins and the Mcf-like protein sequences from *E. gansuensis* var. *inebrians* and *E. typhina* subsp. *poae* 89

Figure 3.5. Mcf phylogenetic trees generated by the (a) branch and bound, and (b) neighbor joining methods 90

Figure 3.6. Expression of *E. typhina* subsp. *poae* Ps1-*mcf in vivo* and in *E. coli* 93

Figure 3.7. Toxicity of the *E. typhina* subsp. *poae* Ps1-Mcf protein to black cutworms 94

Figure 4.1. Reaction catalyzed by salicylate hydroxylase (E.C.1.14.12.1) 105

Figure 4.2. Expression of *E. festucae* Rose City salicylate hydroxylase *in vivo* 114

Figure 4.3. Gene structure of the *E. festucae* Rose City salicylate hydroxylase gene, functional transcript, and alternatively spliced transcript variants 117

LIST OF ILLUSTRATIONS (cont.)

Page

Figure 4.4. Nucleic acid alignment of the experimentally determined *E. festucae* RC salicylate hydroxylase gene and functional transcript sequences in this study, and the five ESTs of transcript variants in the *E. festucae* isolate of *F. pratensis* obtained from NCBI (accession numbers are displayed) 118

Figure 4.5. Colorimetric assay to confirm the heterologous expression of *E. festucae* Rose City salicylate hydroxylase 123

Figure 4.6. Expression of *E. festucae* Rose City salicylate hydroxylase in two *E. coli* expression strains 125

Figure 4.7. Amino acid alignment of salicylate hydroxylase sequences from *E. festucae* Rose City, *U. maydis* Shy1 (XP_761377), and *P. putida* NahG (P23262) 129

Chapter 1

Introduction

Chapter 1

Introduction

Plant-fungi symbiosis

Symbiosis, originating from the Greek word sumbiosis, is widely used to describe a situation where two organisms live together (de Bary, 1879). In that vein, nearly all plants researched in their native environment engage in some form of symbiosis with fungi (Rodriguez and Redman, 2008). There are two main categories of fungal symbiont that are typically observed in association with plants. The key difference between them lies in the location where they reside. Fungal endophytes are characterized as localized entirely within the plant tissues and may also be found in the roots, stems, and/or leaves. Mycorrhizal fungi, on the other hand, are found exclusively in roots, although they stretch out into the rhizosphere (Rodriguez et al., 2004). The nomenclature for mycorrhizae has been updated in recent years to distinguish those that colonize the thalli and shoot systems (paramycorrhizae) from the well-known eumycorrhizal colonizers of the root systems (Strullu-Derrien and Strullu, 2007).

There are vast differences in the documented lifestyles of fungal symbionts. Mutualism, commensalism, and parasitism are the three well-known interactions that span the extremes of symbiont-host relationship. Mutualistic associations are described as beneficial for both parties involved, in contrast to fungi exhibiting parasitic lifestyles that usually affect host fitness antagonistically. Commensal fungi typically do not impact host organisms in any positive or negative manner (Lewis, 1985). In mutualism, the fungal symbiont's nutrient requirement is met without the plant triggering its defense mechanisms. Conversely, the plant gains from this interaction by way of fitness benefits that translates to stress tolerance, higher rates of growth, and/or nutrient gain (Rodriguez et al. 2004). Hence, a state of sophisticated equilibrium is maintained (Kogel et al. 2006).

One of the most intensively studied areas of mutualism is the interaction between plant and arbuscular mycorrhiza fungi of the monophyletic Glomeromycota phylum (Schüssler et al., 2001). Documented to have begun some 460 million years ago, the alliance between the members of Glomeromycota and early plants is conventionally acknowledged to have had played a major role in plant terrestrialisation (Simon et al., 1993; Redecker et al., 2000; Wang et al., 2003). Recent findings, in light of increased taxon sampling, have further supported the crucial contribution of symbiotic fungi towards plant colonization of land. The reported mutualism between *Endogone*-like fungi in the subdivision Mucoromycotina and ancient plant lineages has also highlighted the contribution of a wider repertoire of fungal symbionts, than previously thought, towards plant terrestrialisation (Bidartondo et al., 2011; Desirò et al., 2013; Strullu-Derrien et al., 2014).

Studies on the fungal endophytes of turfgrasses have been documented since the 19th century (Schardl et al., 2004). The *Epichloë* and *Neotyphodium* fungi are mutualistic symbionts of cool-season turfgrasses (White, 1988) and are well documented to impart a wide range of benefits to their plant hosts (Tadych et al., 2014). The fungal endophyte-turfgrass mutualism is an active area of investigation.

Significance of turfgrass

The turfgrass industry in the United States is an important part of the economy. In a comprehensive study based on data from 2002, the US turfgrass industry was estimated to contribute over 70 billion dollars to the federal economy in addition to the creation of close to 823,000 jobs nationwide (Haydu et al., 2008). In the United States alone, the area covered with managed turf exceeds 50 million acres (National Turfgrass Research Initiative, 2003). These comprise several varieties of cool-season turfgrasses grown in more than 60 million lawns and nearly 16,000 golf courses. Cool-season turfgrasses include those generally classified as fine

fescues such as Chewings (*Festuca rubra* L. subsp. *fallax* Thuill., and *F. rubra* L. var. *commutata* Gaud.), hard fescue (*F. longifolia* auct. non Thuill.), sheep fescue (*F. ovina* L.), strong creeping (*F. rubra* L. subsp. *rubra*), and slender creeping (*F. rubra* L. subsp. *trichophylla*, and *F. rubra* subsp. *littoralis* Vassey) red fescues. Also in this category are tall fescue (*F. arundinacea*), Kentucky bluegrass (*Poa pratensis*), bentgrass (*Agrostis* spp.), and perennial ryegrass (*Lolium perenne*).

The significance of grasses is manifold in that they provide both aesthetic and monetary values. This is evidenced by their importance in the maintenance of recreational parks and athletic fields. Additionally, these grasses provide an added beauty to the environment and increase the worth of residential and commercial properties. Turfgrasses are also essential from an environmental perspective because they help to reduce soil erosion, regulate soil temperatures, and capture pollutants (National Turfgrass Research Initiative, 2003).

The strong creeping red fescue

The work described in this thesis is focused on strong creeping red fescue, *F. rubra* L. subsp. *rubra*, native to Europe. This perennial cool-season turf species is reported to have tolerance to acidic, infertile, and droughty soils. It is valued for its moderate shade requirements, and low maintenance (Ruemmele et al., 2003). Additionally, the utility of this species for phytoremediation of copper and manganese in metal-contaminated soil was recently demonstrated (Padmavathiamma and Li, 2009).

Turfgrass-endophyte interaction

On the basis of phylogenetic diversity, taxonomy, plant hosts, and ecological roles, endophytic fungi fall into two distinct categories i.e. clavicipitaceous and non-clavicipitaceous (Rodriguez et al., 2009). The known hosts of clavicipitaceous endophytes are some grasses while the non-

clavicipitaceous endophytes infect non-vascular plants, ferns, conifers, and angiosperms (Rodriguez et al., 2009). Clavicipitaceous endophytes (also known as Class 1 endophytes) are further divided into three groups representing species that have symptoms and pathogenesis (Type I), mixed interaction (Type II), and no symptoms (Type III) (White, 1988).

Class 1 endophytes were first reported in the genus *Lolium* in late 19th century Europe (Rodriguez et al., 2009). Three different publications at that time had documented a presumed relationship between the harmful effects on animals that had eaten the infected grass. Almost 80 years later, the earlier observations were validated when Bacon et al. (1977) connected the phenomenon of summer syndrome toxicosis in livestock foraging tall fescue plants to the endophyte *Neotyphodium coenophialum*, now known as *Epichloë coenophiala* (Leuchtmann et al., 2014).

Recent data indicate that the Clavicipitaceous endophytes originated from an animal pathogen with insect-parasitic ancestors, and through several inter-kingdom host jumps became successful grass symbionts (Spatafora et al., 2007). The grass symbionts such as *E. coenophiala* are believed to have initially been associated with free-living insect parasites and later evolved into distinct endophytic and epibiotic plant biotrophic types as the ability to colonize plant hosts without the need for intermediary organisms developed (Rodriguez et al., 2009). The evolution of endophytes from insect pathogens, instead of being derived from plant pathogens, is postulated as an important explanation as to why these symbionts lack the virulence to damage plants which host them. Therefore, the host plants possibly do not utilize their defense mechanisms to curb the endophytic invasion (Spatafora et al., 2007; Rodriguez et al., 2009).

One of the most widely studied and beneficial endophytic symbionts of cool-season grasses are in

the genus *Epichloë* (Schardl et al., 2004). Type III endophytes are some *Epichloë* species that are asexual (Clay and Schardl, 2002). These species form asymptomatic infections within the intercellular spaces of the vegetative tissues (leaf sheath, culms, and inflorescences), ovules, and seeds (Philipson and Christey, 1986). The propagation of these endophytes is vertically via seeds as they grow on infected mother plants. However, previous studies have reported the natural occurrence of mycelium with conidial formation on the surface of leaf blades in different species of grasses, which implied that seed transmission may not be the sole form of the asexual endophytic fungi proliferation (White et al., 1996; Moy et al., 2000). This observation was recently demonstrated to result from the plant-to-plant horizontal transmission of *Epichloë* conidia (Tadych et al., 2012; reviewed in Tadych et al., 2014).

An interesting characteristic of the Type II grass symbionts is they can exhibit their sexual cycle (by producing sexual stromata) on one tiller and asexual cycle on another tiller of the same plant. The observation made in the 1930s linked *Neotyphodium*, the asexual lifecycle of the endosymbionts with their sexual state (*Epichloë*) (Sampson, 1933; Sampson, 1935). In a recent development, the nomenclature of the endophytes displaying asexual and sexual cycle has been standardized to conform to the one species with a single name description. Hence, the genus *Neotyphodium* has been retired and all fungi formerly in the genus are currently classified as *Epichloë* spp. (Leuchtmann et al., 2014)

In the three decades since Bacon et al.'s initial findings (1977), our knowledge of the *Epichloë* endophytic symbiosis with its grass host has been greatly enriched (Scott, 2001; Clay and Schardl, 2002; Johnson et al., 2003; Rodriguez et al., 2004; Spatafora et al., 2007; Rodriguez and Redman, 2008; Tadych et al., 2012). Two of the best-studied symbiotic interactions are the tall fescue-*E. coenophiala*, and the perennial ryegrass-*E. lolii* interactions (Schardl and Phillips, 1997). Based on these associations, endophytes have been reported to provide a wide range of

benefits to cool-season turf grasses. Among the advantages given by the fungal symbiont to its grass host are enhanced tillering and root growth in addition to increased drought tolerance (Arechavaleta et al., 1989; Malinowski and Belesky, 2000), a range of protective benefits against nematodes (Kimmons et al., 1990), insect pests (Rowan and Latch, 1994; Funk et al., 1983; Dahlman et al., 1991; Murphy et al., 1993), and mammalian herbivory (Bacon et al., 1977; Raisbeck et al., 1991).

In the fine fescues infected with *E. festucae*, protection against insect feeding (Funk et al., 1985; Saha et al., 1987; Yue et al., 2000) has been described. Remarkably, resistance to dollar spot disease caused by the fungal pathogen *Sclerotinia homoeocarpa* (Clarke et al., 2006) and the *Laetisaria fusiformis* red thread infection (Bonos et al., 2005) have also been reported. Dollar spot infection of turf spp. is a destructive and prevalent disease of turfgrasses that is observed worldwide (Walsh et al., 1999). It is also the most expensive turf disease to treat due to its persistent nature of infection (Goodman and Burpee, 1991). Because resistances to diseases have not been detected in other cool-season grass-endophyte interaction to date (Rodriguez et al., 2009), the fine fescue - *E. festucae* relationship is an attractive and important system to examine.

Genome sequence of *Epichloë*

The protection against herbivory and insects that are seen in turf spp. infected with *Epichloë* endophytes is due to the presence of fungal alkaloids (Tanaka et al., 2005; Spiering et al., 2005; Schardl, 2001; Schardl et al., 2012). The well-studied secondary metabolites include indole-diterpenes, aminopyrrolizidines (lolines), pyrrolopyrazine (peramine), and ergot alkaloids (Lyons et al., 1987; Bush et al., 1993; Bush et al., 1997; Schardl et al., 2012). In a recent effort to gain deeper insight into the alkaloid chemicals produced by the *Epichloë* fungal endophytes, Schardl et al. (2013) had undertaken the mammoth effort to sequence 18 species in the genus. These were *E. amarillans* 57, *E. amarillans* 4668, *E. aotearoae, E. baconii, E. brachyelytri, E. bromicola*

AL04262*, E. bromicola* AL0434, *E. coenophiala, E. elymi, E. festucae* 2368, *E. festucae* Fl1, *E. gansuensis, E. gansuensis* var. *inebrians*, *E. glyceriae, E. mollis, E. sylvatica, E. typhina* subsp. *poae* 5809, and *E. typhina* subsp. *poae* E8.

In the report that describes the genomes of the *Epichloë* spp., Schardl et al. (2013) stated that the alkaloid genes formed clusters that are close to AT-rich repeat blocks that originated from transposable elements. They further described that the specific location in the genomes seemed to imply movement and transfer of the alkaloid genes, in addition to being selected for diversification. Since the same pattern was not the norm among other metabolism loci in the genome or the overall genomes themselves, the authors proposed that there was a close relationship between the selection pressure for alkaloid diversification, niche adaptation, and the endophytes' evolutionary history as endosymbionts of cool-season grasses.

Gene expression studies

Available literature examining the basic molecular biology of turfgrass-endophyte interactions is limited but growing, considering the interest in development of endophyte-enhanced turfgrass cultivars. An early study attempted to detect the differential expression of specific genes involved in the tall fescue – *E. coenophiala* symbiosis by utilizing suppression subtractive hybridization (Johnson et al., 2004). The approach yielded information on 29 plant genes that varied in their expression levels in the endophyte-infected and endophyte-free plants. Half of these genes corresponded to known genes deposited in GenBank, and several were identified as having plant defense and stress tolerance functions. This effort underscored the assumption that while both plant host and fungal symbiont are actively involved in the interaction, the endophyte may also engage in the repression of certain plant defense genes.

A study by Zhang et al. (2011) discussed the detection of a pathogenesis-related class 10 (PR-10)

protein in *E. lolii*-infected *L. perenne*, but not in endophyte-free perennial ryegrass or perennial ryegrass-infected with an abnormally growing variant of *E. lolii*. Additionally, a fungal Cu/Zn superoxide dismutase was clearly observed in SDS gels from the protein sample of the endophyte-infected plant although the fungal endophyte content is less than 0.2% of the plant total biomass. The authors wrote that presence of the PR-10 protein was suggestive of a highly restricted host defense in functional and compatible *E. festucae*-perennial ryegrass symbioses. Furthermore they postulated that the role of the fungal Cu/Zn superoxide dismutase was to protect the endophyte from reactive oxygen species present in its environment.

In another recent report also of *E. lolii-L. perenne* interaction using suppression subtractive hybridization (Khan et al., 2010), expressed sequence tag (EST) libraries were generated, and differentially expressed genes in endophyte-infected ryegrass relative to endophyte-free ryegrass were compared. Their study revealed that plant genes involved in carbohydrate metabolism, and photosynthesis were down-regulated during the symbiotic endophyte-ryegrass interaction. Up-regulated ryegrass genes were those involved in cellular protein transport and protein synthesis and turnover. The identification of over 20 *E. lolii* symbiosis-related expressed transcripts were also made. The endophyte transcripts were primarily implicated in endophyte secondary metabolism, while some were found to share similarity to genes involved in pathogenesis.

High-throughput sequencing has entirely revolutionized the field of biological sciences in the last several years. Increasing number of investigators now employ high-throughput sequencing in their system of choice. This is undoubtedly due to the continued accessibility and affordability of the technology to be used as an insightful tool in most areas of study (Faino and Thomma, 2014). The dawn of RNAseq has allowed the exploration of differential gene expression in organisms via high-throughput sequencing transcriptome profiling (Wang et al., 2009). Studies encompassing whole genomes have also directly benefitted from the rise of the DNA sequencing

technology (Koboldt et al., 2013). The first study that took advantage of the sequencing technology to provide new insights into the molecular basis of turfgrass-*E. festucae* symbiosis was reported by in 2010 by Eaton and colleagues (reviewed in Cox et al., 2011; Eaton et al., 2011).

Eaton et al. (2010) used RNAseq to compare gene expression in perennial ryegrass (*L. perenne*) infected with a mutant *E. festucae.* The mutant endophyte's fungal stress- activated mitogen-activated protein kinase gene had been deleted (ΔsakA), and the mutation resulted in a pathogen-like infection of the plant. Interestingly, they detected higher levels of hydrogen peroxide, a known reactive oxygen species in the ΔsakA *E. festucae* mutant infected-plant compared to the wild-type plant. They reported 894 up-regulated genes including those involved in pathogen defense and plant hormones, and higher transposase activity in the transcriptome of the plant infected with the mutant endophyte relative to the plant infected with the wild-type endophyte. Additionally, 308 genes were down-regulated in the ΔsakA *E. festucae* infected-plant when compared with the wild-type sample. Moreover they reported that a large number of the detected fungal genes were specific to *Epichloë*. The study also documented a large number of differentially expressed genes involved in antioxidant enzymes, protein degradation, ethylene biosynthesis, auxin signaling, fungal secondary metabolism, disease resistance, and nitrogen metabolism.

Goals of this thesis

As described earlier, there has been longstanding interest in the relationship between turfgrasses and their *Epichloë* endosymbionts, although it has primarily focused on understanding the fungal alkaloid profiles and the effect of the endophytes on host plant physiology under abiotic and/or biotic stress. The few studies that have attempted to probe the underlying molecular mechanisms driving the symbiosis between cool-season turfgrasses have used the *Epichloë*-infected tall fescue

and perennial ryegrass systems. None to date have investigated the mutualistic symbiosis displayed in the economically important fine fescues' interaction with the fungal endophytes. This is astonishing especially because the *E. festucae*-fine fescue is the only reported *Epichloë*-turfgrass symbiosis to date to confer protection of the host plant against the devastating and damaging dollar spot, and red thread fungal diseases (Clarke et al., 2006; Bonos et al., 2005).

The goal of this thesis is to use SOLiD-SAGE transcriptome analysis to identify candidate genes that may be important in the endophyte-grass symbiosis. High-throughput sequencing is used as a tool to investigate the overall intrinsic processes driving the *Epichloë*-turfgrass interaction and is expected to help achieve the objective.

In the next chapters, findings from three studies are discussed. First, results from a transcriptome-wide quantitative differential gene expression study between clonally propagated strong creeping red fescue infected with *E. festucae* vis-à-vis endophyte-free red fescue are described. Second, the discovery of an intriguing case of horizontal gene transfer into *Epichloë* spp. and the function of the foreign gene in an *Epichloë* sp. is reported. Finally, the salicylate hydroxylase gene is examined in *E. festucae* and its characteristics are detailed. The findings discussed in this thesis contribute to the knowledge on *F. rubra–E. festucae* symbiosis specifically, and the beneficial fungal endosymbiont-plant defensive mutualism in general.

References

1. Arechavaleta M, Bacon CW, Hoveland CS, Radcliffe DE (1989) Effect of the tall fescue endophyte on plant response to environmental stress. Agronomy Journal, 81:83-90.

2. Bacon CW, Porter JK, Robbins JD, Luttrell ES (1977) *Epichloë typhina* from toxic tall fescue grasses. Applied Environmental Microbiology, 34:576-581.

3. Bidartondo MI, Read DJ, Trappe JM, Merckx V, Ligrone R, Duckett JG (2011) The dawn of symbiosis between plants and fungi. Biology Letters, 7:574-577.

4. Bonos SA, Wilson MM, Meyer WA, Funk CR (2005) Suppression of red thread in fine fescues through endophyte-mediated resistance. Applied Turfgrass Science, DOI:10.1094/ATS- 2005-0725- 01-RS.

5. Bush LP, Fannin FF, Siegel MR, Dahlman DL, Burton HR (1993) Chemistry, occurrence and biological effects of saturated pyrrolizidine alkaloids associated with endophyte-grass interactions. Agriculture, Ecosystems & Environment, 44:81-102.

6. Bush LP, Wilkinson HH, Schardl CL (1997) Bioprotective alkaloids of grass-fungal endophyte symbioses. Plant Physiology, 114:1.

7. Clarke BB, White Jr JF, Hurley RH, Torres MS, Sun S, Huff DR (2006) Endophyte-mediated suppression of dollar spot disease in fine fescues. Plant Disease, 90:994–998.

8. Clay K, Schardl CL (2002) Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. American Naturalist, 160:99–127.

9. Cox MP, Eaton CJ, Scott DB (2010) Exploring molecular signaling in plant-fungal symbioses using high throughput RNA sequencing. Plant Signaling & Behavior, 5:1353.

10. Dahlman DL, Eichenseer H, Siegel MR (1991) Chemical perspectives on endophyte-grass interactions and their implications to insect herbivory. Pages 227-252 in: Microbial Mediation of Plant-Herbivore Interactions. Barbosa P, Kirschik L, Jones E, eds. John Wiley & Sons, New York.

11. de Bary A (1879) Die Erschenung Symbiose. Pages 1-30 in: Vortrag auf der Versammlung der Naturforscher und Artze zu Cassel, Strassburg. Trubner KJ, ed. Reviewed in: Rodriguez RJ, Redman RS, Henson JM (2004) The role of fungal symbioses in the adaptation of plants to high stress environments. Mitigation and Adaptation Strategies for Global Change, 9:261–272.

12. Desirò A, Duckett JG, Pressel S, Villarreal JC, Bidartondo MI (2013). Fungal symbioses in hornworts: a chequered history. Proceedings of the Royal Society B, 280:1759.

13. Eaton CJ, Cox MP, Ambrose B, Becker M, Hesse U, Schardl CL, Scott B (2010) Disruption of signaling in a fungal-grass symbiosis leads to pathogenesis. Plant Physiology, 153:1780-1794.

14. Eaton CJ, Cox MP, Scott B (2011) What triggers grass endophytes to switch from mutualism to pathogenism? Plant Science, 180:190-195.

15. Funk CR, Halisky PM, Ahmad S, Hurley RH (1985) How endophytes modify turfgrass performance and response to insect pests in turfgrass breeding and evaluation trials. Pages 127-145 in: Proc. Int. Turf. Res. Conf., 5th. F Lemaire, ed. Avignon, 1–5 July. Versailles, France. INRA Publ.

16. Funk CR, Halisky PM, Johnson MC, Siegel MR, Stewart AV (1983) An endophytic fungus and resistance to sod webworms: association in *Lolium perenne* L. Biotechnology, 1:189–91.

17. Goodman DM, Burpee LL (1991) Biological control of dollar spot disease of creeping bentgrass. Phytopathology, 81:1438–1446.

18. Faino L, Thomma BP (2014) Get your high-quality low-cost genome sequence. Trends in Plant Science, 19:288-291.

19. Haydu JJ, Hodges AW, Hall CR (2008) Estimating the economic impact of the US golf course industry: Challenges and solutions. HortScience, 43:759-763.

20. Johnson LJ, Johnson RD, Schardl CL, Panaccione DG (2003) Identification of differentially expressed genes in the mutualistic association of tall fescue with *Neotyphodium coenophialum*. Physiological and Molecular Plant Pathology, 63:305–317.

21. Khan A, Bassett S, Voisey C, Gaborit C, Johnson L, Christensen M, McCulloch A, Bryan G, Johnson R (2010) Gene expression profiling of the endophytic fungus *Neotyphodium lolii* in association with its host plant perennial ryegrass. Australasian Plant Pathology, 39:467-476.

22. Kimmons CA, Gwinn KD, Bernard EC (1990) Nematode reproduction on endophyte-infected and endophyte-free tall fescue. Plant Disease, 74:757-761.

23. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. Cell, 155:27-38.

24. Kogel K-H, Franken P, Huckelhoven R (2006) Endophyte or parasite – what decides? Current Opinion in Plant Biology, 9:358–363.

25. Leuchtmann A, Bacon CW, Schardl CL, White Jr JF, Tadych M (2014) Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. Mycologia, 106: 202-215.

26. Lewis DH (1985) Symbiosis and mutualism: crisp concepts and soggy semantics. In: Boucher DH, ed. The biology of mutualism. London: Croom Helm Ltd, 29–39.

27. Lyons PC, Plattner RD, Bacon CW (1986) Occurrence of peptide and clavinet ergot alkaloids in tall fescue grass. Science, 232:487–489.

28. Malinowski DP, Belesky DP (2000) Adaptations of endophyte-infected cool season grasses to environmental stresses: Mechanisms of drought and mineral stress tolerance.

Crop Science, 40:923–40.

29. Moy M, Belanger F, Duncan R, Freehoff A, Leary C, Meyer W, Sullivan R, White Jr JF (2000) Identification of epiphyllous mycelial nets on leaves of grasses infected by clavicipitaceous endophytes. Symbiosis, 28: 291-302.

30. Murphy JA, Sun S, Betts LL (1993) Endophyte-enhanced resistance to billbug (*Coleptera*: *Curculionidae*), sod webworm (*Lepidoptera*: *Pyralidae*), and white grub (*Coleoptera*: *Scarabaeidae*) in tall fescue. Environmental Entomology, 22:699–703.

31. National Turfgrass Research Initiative (2003) http://www.ntep.org/pdf/turfinitiative.pdf.

32. Padmavathiamma P K, Li LY (2009) Phytoremediation of metal-contaminated soil in temperate humid regions of British Columbia, Canada. International Journal of Phytoremediation, 11:575-590.

33. Philipson MN, Christey MC (1986) The relationship of host and endophyte during flowering seed formation and germination of *Lolium perenne*. New Zealand Journal of Botany, 24:125–134.

34. Raisbeck MF, Rottinghaus GE, Kendall JD (1991) Effects of naturally occurring mycotoxins on ruminants. Pages 647-677 in: Mycotoxins and Animal Foods. JE Smith, Henderson RS, eds. CRC Press, Boca Raton, FL.

35. Redecker D, Kodner R, Graham LE (2000) Glomalean fungi from the Ordovician. Science, 289:1920-1921.

36. Rodriguez R, Redman R (2008) More than 400 million years of evolution and some plants still can't make it on their own: plant stress tolerance via fungal symbiosis. Journal of Experimental Botany, 59: 1109–1114.

37. Rodriguez RJ, Redman RS, Henson JM (2004) The role of fungal symbioses in the adaptation of plants to high stress environments. Mitigation and Adaptation Strategies for Global Change, 9:261–272.

38. Rodriguez RJ, White JF Jr, Arnold AE, Redman RS (2009) Fungal endophytes: diversity and functional roles. New Phytologist, 182:314–330.

39. Rowan DD, Latch GCM (1994) Utilization of endophyte-infected perennial ryegrasses for increased insect resistance. Pages 169-183 in: Biotechnology of Endophytic Fungi of Grasses. Bacon CW, White JF Jr, eds. CRC Press, Boca Raton,FL.

40. Ruemmele BA, Wipff JK, Brilman L, Hignight KW (2003) Fine-leaved *Festuca* species. Pages 129-174 in: Turfgrass Biology, Genetics and Breeding. Casler MD, Duncan RR, eds. New York: John Wiley & Sons.

41. Saha DC, Johnson-Cicalese JM, Halisky PM, van Heemstra MI, Funk CR (1987) Occurrence and significance of endophytic fungi in the fine fescues. Plant Disease, 71:1021–24.

42. Sampson K (1933) The systemic infection of grasses by *Epichloe typhina* (Pers.) Tul. Transactions of the British Mycological Society, 18:30–47.

43. Sampson K (1935) The presence and absence of an endophytic fungus in *Lolium temulentum* and *L. perenne*. Transactions of the British Mycological Society, 19:337–43.

44. Schardl CL, Leuchtmann A, Spiering MJ (2004) Symbioses of grasses with seedborne fungal endophytes. Annual Review of Plant Biology, 55:315–40.

45. Schardl CL (2001) *Epichloë festucae* and related mutualistic symbionts of grasses. Fungal Genetics and Biology, 33:69–82.

46. Schardl CL, Phillips TD (1997) Protective grass endophytes: Where are they from and where are they going? Plant Disease, 81:430-438.

47. Schardl CL, Young CA, Faulkner JR, Florea S, Pan J (2012) Chemotypic diversity of epichloae, fungal symbionts of grasses. Fungal Ecology, 5:331-344.

48. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk JW, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J, Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach J, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z. (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the *Clavicipitaceae* reveals dynamics of alkaloid loci. PLoS Genetics, 9(2):e1003323.

49. Schüssler A, Schwarzott D, Walker C (2001) A new fungal phylum, the Glomeromycota: phylogeny and evolution. Mycological Research, 105:1413-1421.

50. Scott B (2001) *Epichloë* endophytes: fungal symbionts of grasses. Current Opinion in Microbiology, 4:393–398.

51. Simon L, Bousquet J, Lévesque RC, Lalonde M (1993) Origin and diversification of endomycorrhizal fungi and coincidence with vascular land plants. *Nature,* 363*:*67–69.

52. Spatafora JW, Sung G-H, Sung J-M, Hywel-Jones NL, White JF Jr (2007) Phylogenetic evidence for an animal pathogen origin of ergot and the grass endophytes. Molecular Ecology, 16:1701–1711.

53. Spiering MJ, Moon CD, Wilkinson HH, Schardl CL (2005) Gene clusters for insecticidal loline alkaloids in the grass-endophytic fungus *Neotyphodium uncinatum*. Genetics, 169:1403–1414.

54. Strullu-Derrien C, Strullu DG (2007) Mycorrhization of fossil and living plants. Comptes Rendus Palevol, 6:483-494.

55. Strullu-Derrien C, Kenrick P, Pressel S, Duckett JG, Rioult JP, Strullu DG (2014) Fungal associations in *Horneophyton ligneri* from the Rhynie Chert (c. 407 million year old) closely resemble those in extant lower land plants: novel insights into ancestral plant–fungus symbioses. New Phytologist, doi:10.1111/nph.12805.

56. Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B (2005) A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. Molecular Microbiology, 57:1036–1050.

57. Tadych M, Ambrose KV, Bergen MS, Belanger FC, White Jr JF (2012) Taxonomic placement of *Epichloë poae* sp. nov. and horizontal dissemination to seedlings via conidia. Fungal Diversity, 54:117-131.

58. Tadych M, Bergen MS, White JF (2014) *Epichloë* spp. associated with grasses: new insights on life cycles, dissemination and evolution. Mycologia, 106: 181-201.

59. Walsh B, Ikeda SS, Boland GJ (1999) Biology and management of dollar spot (*Sclerotinia homoeocarpa*); an important disease of turfgrass. HortScience, 34:13-21.

60. Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: A revolutionary tool for transcriptomics. Nature Review Genetics, 10:57–63.

61. Wang B, Yeun LH, Xue JY, Liu Y, Ané JM, Qiu YL (2010) Presence of three mycorrhizal genes in the common ancestor of land plants suggests a key role of mycorrhizas in the colonization of land by plants. New Phytologist, 186:514-525.

62. White JF Jr (1988) Endophyte-host associations in forage grasses. IX. A proposal concerning origin and evolution. Mycologia, 80: 442-446.

63. White JF Jr, Martin TI, Cabral D (1996) Endophyte-host associations in grasses. XXII. Conidia formation by *Acremonium* endophytes on the phylloplanes of *Agrostis hiemalis* and *Poa rigidifolia*. Mycologia, 88:174-178.

64. White JF Jr, Torres MS (2010) Is plant endophyte-mediated defensive mutualism the result of oxidative stress production? Physiologia Plantarum, 138:440-446.

65. Yue Q, Gianfagna TJ, Meyer WA (2000) Alkaloid production and chinch bug resistance in endophyte-inoculated Chewings and strong creeping red fescues. Journal of Chemical Ecology, 26:279–292.

66. Zhang N, Zhang S, Borchert S, Richardson K, Schmid J (2011) High levels of a fungal superoxide dismutase and increased concentration of a PR-10 plant protein in associations between the endophytic fungus *Neotyphodium lolii* and ryegrass. Molecular Plant-Microbe Interactions, 24:984-992.

Chapter 2

SOLiD-SAGE of endophyte-infected red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins

Chapter 2

SOLiD-SAGE of endophyte-infected red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins

Introduction

An interaction that is of broad significance in turf is the symbiosis between *Epichloë* endophytic fungi and cool season grasses (Schardl, 2001). The endophyte-grass interaction is valued commercially and is also a promising model for studies aimed at understanding symbiotic associations in general. It is well established that the *Epichloë* fungal endophytes of grasses confer numerous benefits to their hosts (Rodriguez et al., 2009; Kuldau and Bacon, 2008; Schardl et al., 2009). The endophytic fungal hyphae ramify within the intercellular spaces of the aerial plant parts, in particular the leaf sheaths (Hinton and Bacon, 1985). They grow through the host apoplastic spaces by a novel intercalary hyphal extension process, rather than exclusively by hyphal tip extension (Christensen et al., 2008). They do not invade the plant cells and must therefore obtain all their carbon and nitrogen compounds from the apoplastic space. However, the host must in some way sense the presence of the endophyte since studies have shown that presence of *Epichloë* spp. endophytes results in changes of host gene expression in tall fescue (*Festuca arundinacea*, syn. *Lolium arundinaceum*) (Johnson et al., 2003; Dinkins et al., 2010) and perennial ryegrass (*L. perenne*) (Eaton et al., 2010). Several studies indicate endophyte infection can result in increased plant vigor and confer tolerance to abiotic stress, unrelated to the reduction in herbivory (Kuldau and Bacon, 2008). The physiological mechanisms that produce these effects are not understood, but it seems likely some changes in host gene expression may be a factor.

Another well-established effect of some grass-endophyte symbioses is disease resistance, but this trait is apparently unique to endophyte-infected fine fescues since resistance to fungal pathogens is not observed in *Epichloë* endophyte infection of other grass species (Rodriguez et al., 2009).

Individuals of these species are often naturally infected with the fungal endophyte *E. festucae* (Schardl, 2001; Scott et al., 2012). Many current fine fescue cultivars are infected with endophytes because endophyte infection is generally desirable because of the insect and disease resistance conferred on the host grass. In field evaluations, endophyte-infected fine fescues exhibited enhanced resistance to the fungal diseases dollar spot and red thread caused by *Sclerotinia homoeocarpa* and *Laetisaria fuciformis*, respectively (Clarke et al., 2006; Bonos et al., 2005), but there is no information as to the mechanism of the observed disease resistance seen in this symbiotic association. The endophyte-mediated disease resistance is agronomically important, since it reduces the fungicide requirements for these low maintenance grasses.

New high-throughput sequencing systems have revolutionized genome sequencing as well as expression analysis (Rusk and Kiermer, 2008; Lister et al., 2009). Now it is possible to quantitatively compare transcript abundance among many samples at relatively low cost. Such new approaches can considerably expand the information generated from previous approaches regarding the effect of endophyte infection on host gene expression.

In this chapter, the use of SOLiD-SAGE, a high-throughput adaptation of Serial Analysis of Gene Expression (SAGE) (Matsumura et al., 2010) using Life Technologies' Sequencing by Oligonucleotide Ligation and Detection (SOLiD) platform, to quantitatively compare transcript abundance in endophyte-free and endophyte-infected strong creeping red fescue (*Festuca rubra* L. subsp. *rubra*) is described. This analysis revealed hundreds of plant genes involved in many different physiological processes whose expression levels were affected by the presence of the fungal endophyte. The results also showed many of the highly expressed fungal endophyte transcripts are comprised of transcripts for secreted proteins. One of these abundant transcripts encodes an antifungal protein that appears to be unique to *Epichloë* endophytes infecting *F. rubra* and *Achnatherum inebrians* (drunken horse grass). It is a candidate for involvement in the

observed endophyte-mediated disease resistance in *F. rubra*.

The work discussed here is published in Ambrose KV, Belanger FC (2012) SOLiD-SAGE of endophyte-infected red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins. PLoS ONE, 7:e53214.

Materials and methods

Plant and fungal materials

Strong creeping red fescue plants S1139E- and S1139RC were described previously (Johnson-Cicalese et al., 2000). The endophyte-infected plant S1139RC was generated by inoculating an isolated tiller of the uninfected plant S1139E- with the Rose City isolate of *E. festucae*, which was isolated from an unrelated endophyte-infected strong creeping red fescue (Johnson-Cicalese et al., 2000). The S1139E- and S1139RC plants thus represent endophyte-free and endophyte-infected examples of the same plant genotype. The endophyte status of the plants was confirmed microscopically prior to their use in this study. These plants can be clonally propagated and were maintained in 6-inch pots in a greenhouse.

Plants for RNA isolation for SOLiD-SAGE analysis were grown in a growth chamber set to 16 h light (400 μmol m^{-2} s^{-1}) at 21C, 8 h dark at 19C, and 50-55% relative humidity. Seventy-two individual tillers of each plant type were planted in plastic 6-cell trays (one tiller per cell) and maintained in the growth chamber for 56 days. After 1 week in the growth chamber, the plants were fertilized with 10-3-20 Peat Lite Plant Starter (Everris), and 30 pellets of Osmocote 14-14-14 (Scotts Miracle-Gro) were added to each cell.

E. festucae was isolated from the endophyte-infected S1139RC plant by plating surface-sterilized leaf sheath tissue on potato dextrose agar (Difco Laboratories, Detroit, MI).

RNA and DNA isolation

For RNA for the 454 pyrosequencing, leaf sheath tissue of S1139RC grown in the greenhouse was used. For RNA for the SOLiD-SAGE library preparation, three biological replicates of the inner most leaf sheath tissue of S1139E- and S1139RC were harvested after 56 days in the growth chamber. For isolation of RNA, the fungus was grown in potato dextrose broth for 9 days on a shaker (175 rpm) at room temperature.

For RNA isolation, each 1 g sample was ground to a fine powder with liquid nitrogen and resuspended in 10 mL Tri-Reagent (Sigma-Aldrich, St. Louis, Mo, USA). Debris was removed by centrifugation and supernatant was extracted twice with chloroform. RNA in the aqueous layer was precipitated with isopropanol, and the RNA pellet was washed once with ethanol and dissolved in water. Fungal genomic DNA was extracted from a culture grown in potato dextrose broth for 14 days. The DNA was isolated as previously described (Moy et al., 2002).

454/GS FLX Titanium sequencing

To provide homologous reference sequence datasets for the SOLiD-SAGE analysis we generated transcriptome sequences for single replicates of plant S1139RC and for the Rose City fungal isolate grown in culture by using the Roche 454 pyrosequencing platform. The endophyte-infected plant was chosen for the 454 sequencing since it was a possibility that some plant and/or fungal genes may only be expressed in endophyte-infected tissue. The longer sequences generated by 454 pyrosequencing facilitated gene identification of the shorter 27 bp SOLiD-SAGE tags.

To maximize the transcriptome coverage, the cDNA samples prepared for 454 pyrosequencing were normalized prior to sequencing. cDNA normalization results in an equalization of transcript concentrations in the population so that highly abundant transcripts do not overwelm the subsequent sequencing. In the normalization process the cDNA is denatured and then reassociated.

Duplex-specific nuclease then is used to degrade the double-stranded cDNA fraction formed by abundant transcripts. Samples for 454 pyrosequencing were normalized to obtain sequences for as many different transcripts as possible to aid in the identification of the quantitative SOLiD-SAGE data.

Barcoded cDNA library construction and 454 pyrosequencing were performed by the University of Georgia Genomics Facility. The Evrogen MINT-Universal cDNA synthesis kit (Axxora, LLC, San Diego, CA) was used with a modified oligo(dT) primer (5'-AAGCAGTGGTATCAACGCAGAGTAC$(T)_4$G$(T)_9$C$(T)_{10}$VN-3') for the first strand synthesis. The Evrogen Trimmer kit (Axxora, LLC) was used to normalize the cDNAs. The cDNA samples were prepared for sequencing by following the rapid library preparation method as described in the GS FLX Titanium manual (454 Life Science, Branford, CT). Further processing was done according to the manufacturer's protocols. The cDNAs were combined in a ratio of 3:1 (plant S1139RC:endophyte) and run on 1/4 plate for sequencing. The 454 sequences are available in the GenBank Sequence Read Archive (SRA) database under accession number SRA052297.

***De novo* 454 transcriptome assembly**

Raw 454 reads shorter than 50 nucleotides were removed. The *F. rubra* endophyte-infected plant and the *E. festucae* fungal endophyte 454 reads were each assembled without references into contigs by using SeqMan NGen v3.1 (DNAStar Inc., Madison, WI). Parameters for both assemblies were minimum match size of 21 nucleotides, minimum match percentage of 85%, mismatch penalty of 20, and gap penalty of 30. A BLASTn comparison of all the endophyte-infected *F. rubra* contigs and singletons against the *E. festucae* E2368 whole genome sequence (GenBank accession ADFL00000000) was done to identify those sequences that originated from the endophyte (E-value cutoff of 1e-05).

SOLiD-SAGE sequencing

SOLiD-SAGE libraries were prepared from each of the three biological RNA replicates from plants S1139E- and S1139RC (Figure 2.1). Each library was generated by using a kit (SOLiD SAGE S3100301, Applied Biosystems, Foster City, CA), which was modified to facilitate barcoding of the samples. Instead of adaptor A supplied in the kit, adaptor Abc (5'-GTACGGCCAAGGCGGATGTACGGTACAGCAGCATG-3') was used. Adapter Abc contains a 4-bp overhang (CATG), which complements the *Nla*III digested double stranded cDNA, an *Eco*P15I restriction enzyme recognition site at the 3' end, and a PCR priming site. Oligos for barcode addition were obtained from the SOLiD Small RNA Expression Kit (Applied Biosystems) and were added to the tags by PCR. Barcode addition, emulsion PCR, and SOLiD sequencing of the libraries were performed at the Waksman Institute Genomics Core Facility, Rutgers University, New Brunswick, NJ. The barcoded libraries were combined and sequenced on one-quarter of a slide on an Applied Biosystems SOLiD 4 System.

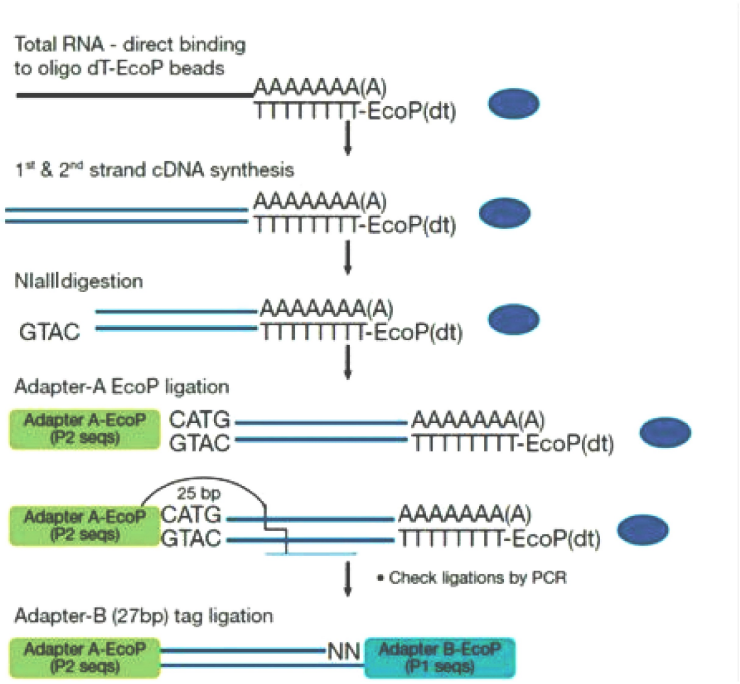


Figure 2.1. Standard SOLiD-SAGE tag construction overview. SOLiD-SAGE tags for nine libraries were generated from total plant RNA of *E. festucae* Rose City-infected, or endophyte-free red fescues following the protocol shown in the diagram. Adapter A supplied in the kit (shown here) was replaced with the modified adapter Abc. Each 27-bp tag represents a transcript in the samples (Image obtained from Life Technologies, Thermo Fisher Scientific).

Computational analysis of SOLiD-SAGE tags

The characteristics of each SAGE library were determined by using the Galaxy bioinformatics suite (http://main.g2.bx.psu.edu) (Goecks et al., 2010; Blankenberg et al., 2010; Giardine et al., 2005). The total number of SAGE tags with the *Nla*III restriction enzyme recognition site in each library were first determined. Those SAGE tags containing the *Nla*III site were then analyzed for tags containing homopolymer A (10 or more nucleotides long) at the 3' end, which is suggestive of SAGE tags containing poly(A)$^+$ tails. The total number of putative poly(A)$^+$-containing SAGE tags was subtracted from the number of SAGE tags containing *Nla*III site in order to determine the number of potential mappable SAGE tags.

Analysis of the SOLiD-SAGE tags was done by using the Applied Biosystems software program (SOLiD-SAGE v1.10) that maps the SAGE tags to a user supplied sequence database and returns the number of times a particular tag is found in each library. For mapping, the tag length was set to 27 bases and the maximum mismatches allowed was set to two. To separate the plant and fungal SAGE tags, they were mapped in two steps to a reference data set consisting of **1**) the S1139RC 454 sequences generated in this study and the *Festuca* and *Lolium* sequences (141,259) downloaded from NCBI or **2**) the *E. festucae* 454 sequences generated in this study, *Epichloë* sequences (57,687) downloaded from NCBI, and the whole genome sequence of *E. festucae* isolate E2368 (GenBank accession ADFL02000000).

To identify those plant tags that originated from differentially expressed transcripts, the raw tag counts were normalized by converting to number of tags per million mapped tags in that library. Statistical significance ($P < 0.05$) of differential gene expression between plants S1139E- and S1139RC was determined from unpaired t tests by using the PRISM 4 program (GraphPad Software, San Diego, CA). For comparing transcript abundance of the fungal transcripts, the raw tag counts in the replicate libraries were converted to percent of mapped tags.

Gene identification of the plant and fungal SAGE tags was by BLAST searches of the corresponding 454 sequences to NCBI databases or to the annotated *E. festucae* genome sequence (Schardl et al., 2013). The program TargetP (http://www.cbs.dtu.dk/services/TargetP/) (Emanuelsson et al., 2007) was used to predict secreted proteins.

GO annotation

GO Slim categorization of the differentially expressed plant transcripts was done by using Blast2GO v.2.5.0 (Conesa et al., 2005; Conesa et al., 2008; Götz et al., 2008; Götz et al., 2011) and QuickGO (Binns et al., 2009; http://www.ebi.ac.uk/QuickGO/GMultiTerm).

RT-PCR

Oligonucleotide primer sequences used for cDNA synthesis and PCR amplification are presented in Table 2.1. First-strand cDNA of 4 µg S1139RC total RNA was synthesized from either 500 ng of oligo$(dT)_{18}$ primer or 2 picomoles of a strand-specific primer by using SuperScript™ III Reverse Transcriptase (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions.

PCR was performed in 20 µL with either 1 µg of fungal genomic DNA, 5 µL *E. festucae*-infected plant cDNA generated from oligo(dT) or 1 µL *E. festucae*-infected plant cDNA generated from a gene-specific primer as templates, 0.25 mM each dNTP, 20 picomoles of each forward and reverse primer (Integrated DNA Technologies, Inc., Coralville, IA), 2 µL of 10X AmpliTaq Buffer and 0.2 µL AmpliTaq Gold DNA polymerase (Life Technologies, Inc., Carlsbad, CA). PCR was done in a GeneAmp 9700 thermocycler (Applied Biosystems, Inc., Foster City, CA). The initial denaturation was conducted at 94°C for 2 min, followed by 35 cycles of 30 s denaturation at 94°C, 30 s annealing at 55°C, and 1 min extension at 72°C, followed by a final extension at 72°C for 10 min. The amplification products were visualized on a 2% agarose gel.

DNA sequencing of *Ef-AFP* splice variant

Forward and reverse primers were designed based on the *E. festucae* genome sequence (Schardl et al., 2013) to amplify the entire region of the *Ef-AFP* gene, and a partial area of the *Ef-AFP* alternatively spliced variant (Table 2.6). The RT-PCR product of the partial *Ef-AFP* alternatively spliced variant was sequenced directly (Genewiz, Inc., South Plainfield, NJ). For each sequencing reaction, a 10 µL aliquot of the PCR product was treated with 2 µL ExoSAP-IT (USB Corp., Cleveland, OH) to remove unincorporated primers and excess dNTPs. The ExoSAP-IT reaction was performed at 37°C for 15 min followed by heating at 80°C for 15 min to inactivate the enzymes. Sequencing was done in both directions.

Phylogenetic analysis

The CLUSTAL-X program (Thompson et al., 1997) was used to align the DNA sequences. For the phylogenetic analyses the sequences were trimmed to include only the regions of sequence overlap for all the sequences in the analysis. The alignments generated by Clustal X were modified manually to minimize gaps. The phylogenetic analyses were performed with the PAUP* program (version 4.0b10 for Macintosh; Swofford, 2002). For both the MCM7 and antifungal protein gene phylogenetic analyses, introns were removed and only the protein coding sequences were used.

The MCM7 phylogenetic analysis was done by using the maximum parsimony full heuristic search option set to random sequence addition, tree-bisection-reconnection (TBR) branch swapping, and Multrees on, with 1,000 bootstrap replications. Gaps were treated as missing data. The *Tuber melanosporum* (Class Pezizomycetes) sequence was designated as the outgroup to root the tree since it is basal to the other species in the tree (Lutzoni et al., 2004). The antifungal protein gene phylogenetic tree was done by using an exhaustive maximum parsimony analysis, which returned a single most parsimonious tree. The tree was midpoint rooted.

Table 2.1. Sequences of oligonucleotide primers used in this study.

Reference accession	Forward primer, 5' – 3'	Reverse primer, 5' – 3'	Amplicon size (bp)
SRR4936	**Amplification of *Ef-AFP***		
91.12929	ATGCAAATCACCGTGGTCGC	CTAATGACACGTGACAGCTC	423 (genomic) 279 (cDNA)
ADFL020	**Amplification of *Ef-AFP* alternatively spliced variant**		
00476	AATACCAGACAAAAGGGTCGC	TCAGTCTAGTTCTTCCCTAGA	195 (genomic) 195 (cDNA)
SRR4936	**Strand specific amplification of *Nc12* sense transcript**		
91.19548	ATTCGCTGGAGAAGACCATG	TAGTCTGGCTAGCAAGAAGG	174
	Strand specific amplification of *Nc12* antisense transcript		
	ATGGTCTTCTCCAGCGAATG	CCTGTTACATACTCGCTCTC	192

	Strand specific amplification of *Ef-AFP* sense transcript		
SRR493691.12929	GACAATCTGATTCCTCTCTTTC	GACGGGCACTTGACGAACG	250
	Strand specific amplification of *Ef-AFP* antisense transcript		
	GTACTTGCATTCGTTTTTGGCT	ACCGTGGTCGCGGTTTTCC	150
	Strand specific amplification of *E. festucae* protease sense transcript		
SRR493691.55163	TGGGCGCCGACGGGCAG	TAAATAACCTATACTATTCTC TATA	250
	Strand specific amplification of *E. festucae* protease antisense transcript		
	ACATAACAACAGCAACAGACTT	GGCGCCGACGGGCAGAA	150
	Strand specific amplification of *F. rubra* metallothionein sense transcript		
HO060295.1	GCGGGAAGATGTACCCTGA	GATGAAGCAATTACAGAGAT ATA	350
	Strand specific amplification of *F. rubra* metallothionein antisense transcript		
	GTCAAACCTGTTTTTCTTAAGTA	GTTTACTTGCTCGCTATGCTA	150

Results

454 transcriptome sequencing

The plant materials used in the study were a strong creeping red fescue not infected with *E. festucae*, designated S1139E-, and the same plant genotype inoculated with the *E. festucae* Rose City strain, designated S1139RC (Johnson-Cicalese et al., 2000). The Rose City isolate was obtained from an endophyte-infected strong creeping red fescue plant. The Rose City isolate of *E. festucae* has been demonstrated to confer insect resistance and fungal pathogen resistance to its host (Clarke et al., 2006; Bonos et al., 2005; Yue et al., 2000).

The characteristics of the 454 transcriptome sequences are summarized in Table 2.2. Over 200,000 total sequences with an overall average length of 307 bp were obtained. The S1139RC 454 sequences, most of which represent plant transcripts, were assembled into 68,817 plant unigenes (plant contigs plus plant singletons). The diploid monocot rice has 41,000 genes (Sterck et al., 2007). *F. rubra* is a hexaploid (Smarda et al., 2007), so a crude estimate of the gene content is 123,000 genes. The 454 sequences therefore are estimated to cover 56% of the total *F. rubra* gene content.

Table 2.2. Characteristics of the 454 sequences.

	***Festuca rubra*, endophyte-infected (S1139RC)**	***Epichloë festucae* (from culture)**
Total raw reads (n)	122,041	79,654
Mean length of raw reads (nt)	309	305
Total nucleotides of raw reads (nt)	37,717626	24,339,130
Post trim total reads (n)	118,862	77,254
Post trim mean length of reads (nt)	316	314
Post trim total nucleotides (nt)	37,592,208	24,245,046
Reads assembled into contigs (n)	67,553	54,638
Mean length of assembled reads (nt)	287	321
Total contigs (n)	20,501	13,381
Contig N50	426	544
Mean length of contigs (nt)	382	470
Plant origin contigs (n)	19,496	NA
Fungal endophyte origin contigs (n)	1,005	NA
Total singletons (n)	51,309	22,616
Mean length of singletons (nt)	278	277
Plant origin singletons (n)	49,321	NA
Fungal endophyte origin singletons (n)	1,988	NA

Differentially expressed host plant genes identified by SOLiD-SAGE sequencing

SAGE (Serial Analysis of Gene Expression) and the improved method SuperSAGE (Velculescu et al., 1995; Matsumura et al., 2003; Matsumura et al., 2005) have been used in numerous studies for transcriptome analysis. SuperSAGE generates 27 bp cDNA sequences, and in SAGE terminology the sequences are called “tags”. SuperSAGE in combination with the high-throughput sequencing capability of the SOLiD sequencing system was used here. SOLiD-SAGE generates 27 bp tags from the most 3’ *Nla*III restriction site (recognition site: 5’-CATG-3’) in each cDNA (Matsumura et al., 2010; Hong et al., 2011). The 4 base *Nla*III restriction site is expected to be frequent and to occur on average every 256 bp ($4^4 = 256$), but if a transcript does not have an *Nla*III site it will not be represented among the tags. The number of times a particular tag is sequenced is directly related to the number of transcripts in the sample (one tag = one transcript). Relative transcript

levels can therefore be determined. An overview of the SOLiD-SAGE tag construction is shown in Figure 2.1.

Triplicate biological samples for each of the plant genotypes were prepared. Triplicate biological samples are critical for evaluation of statistical significance of differential gene expression. The characteristics of the SAGE tags are summarized in Table 2.3. Between 5 and 10 million total tags per sample were obtained. The total tags were filtered to remove any tags that did not contain the *Nla*III cleavage site (CATG) at the 5' end. Such tags must have originated from random ligation of adaptors to the cDNA and cannot be used in a quantitative assessment of transcript abundance. Those tags containing the *Nla*III site were further filtered to remove any that had greater than 10 As at the 3' end. The SAGE tags are generated from the most 3' *Nla*III site and are 27 bp long. If the *Nla*III site is close to the end of the transcript, then some of the resulting tag sequence will contain part of the poly(A)$^+$ region of the cDNA. Such tags will not map to the reference dataset. Since it is impossible to distinguish between As that are truly part of the transcript sequence versus As originating from poly(A)$^+$, an arbitrary number of 10 was chosen.

Table 2.3. Characteristics of the SOLiD-SAGE tags.

	Total tags	**Tags with *Nla*III site**	**Tags after removal of those with > 10 As**	**Plant mapped tags (% of potential mappable)**	**Fungal mapped tags (% of potential mappable)**
S1139E-					
Replicate 1	10,266,193	7,208,225	7,194,891	4,346,031 (60%)	NA
Replicate 2	8,048,623	5,629,653	5,613,295	3,303,965 (59%)	NA
Replicate 3	6,015,959	4,019,984	4,012,492	2,269,416 (56%)	NA
S1139RC					
Replicate 1	5,415,607	3,044,831	3,041,847	1,137,472 (37%)	43,199 (1.4%)
Replicate 2	6,165,530	4,031,091	4,025,153	2,160,008 (54%)	112,294 (2.8%)
Replicate 3	5,736,324	3,364,721	3,361,001	1,471,380 (44%)	71,927 (2.1%)

The number of tags remaining after the two filtering steps was between 3 and 7 million. These tags represent the expected number of tags that could be mapped. Tags that actually mapped to the plant reference dataset ranged from 37% to 60% of the expected number of mapped tags. The difference in the expected number of mappable tags and the actual number of mapped tags is likely due to the lack of a complete *F. rubra* reference dataset. As described above, the *F. rubra* 454 sequences were estimated to cover 56% of the plant transcriptome. Additional *Festuca* and *Lolium* EST sequences from NCBI were included in the mapping database, so the overall coverage of the plant genes could be expected to be higher than 56%, but it is likely that some plant genes were not represented in the mapping database. Even with this limitation to the SOLiD-SAGE approach with a nonmodel system, millions of tags could be mapped and plant genes with differential expression could be identified.

The most abundant host plant genes

The same tag sequence was the most abundant tag in all 6 libraries and was a match to a 454 sequence that was identified through a BLASTx search as a chlorophyll a/b binding protein. For comparison of expression levels between S1139E- and S1139RC the number of tags in each biological replicate was normalized to the total number of tags per million mapped tags (TMM) in that replicate. There was no significant difference in expression level of the chlorophyll a/b binding protein between the two samples. The tag for the chlorophyll a/b binding protein represented a mean of 12,282 and 11,922 TMM of the mapped tags in S1139E- and S1139RC, respectively. Other abundant tags in all the libraries that were not significantly different between the samples were identified as originating from transcripts for metallothionein and ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit protein. These three abundant plant transcripts have all been reported as abundant in many plant EST libraries and previous SAGE analyses (Bausher et al., 2003; Gibbings et al., 2003; Robinson et al., 2004; Jantasuriyat et al., 2005; Pratt

et al., 2005) indicating that the SOLiD-SAGE libraries are a good reflection of plant transcript abundance in the generated samples.

Physiological categories of differentially expressed plant genes

The plant SAGE tags were searched for those having statistically significant ($P < 0.05$) differential expression levels between S1139E- and S1139RC, and 209 such tags were identified. The complete list of differentially expressed plant genes can be found as Supporting Information Table S1A and B in Ambrose and Belanger (2012). Most (182) of the differentially expressed genes could be assigned a protein identification and these genes fell into 31 gene ontology (GO) categories. Twenty-seven of the differentially expressed tags originated from transcripts for as yet uncharacterized proteins. A summary of differentially expressed tags by gene ontology category is presented in Table 2.4. Some of the differentially expressed SAGE tags were SNPs of each other and could be mapped to different 454 sequences that encoded the same protein. Such tags represent alleles or alloalleles of each other. The genes showing differential expression ranged from abundantly expressed (lipid transfer proteins) to those with relatively low expression levels (stress induced hydrophobic proteins). The fold changes of the differentially expressed genes ranged from +7.4 to -7.2.

One GO category of differentially expressed genes that stood out was that of photosynthesis, which comprised 13% of the up-regulated genes in the endophyte-infected plants. The SAGE tags for the up-regulated chlorophyll a/b binding protein and ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit protein are from different members of the gene families for these proteins than the abundant SAGE tags discussed above. Numerous studies have reported increased productivity and photosynthetic rates in endophyte infected plants (Clay and Schardl, 2002), although when under high light the photosynthetic rate in endophyte infected perennial ryegrass was reported to decrease (Spiering et al., 2006). Overall, the presence of the fungal

endophyte resulted in modest changes in expression level for plant genes involved in a wide range of physiological processes.

Previous studies used suppression subtractive hybridization, and microarray analysis to identify plant genes differentially expressed in response to endophyte infection in tall fescue (Johnson et al., 2003; Dinkins et al., 2010). There was little overlap with the previous studies and the differentially expressed plant genes identified here through SOLiD-SAGE. SOLiD-SAGE tags were identified for most of the plant genes reported in the previous two studies to be differentially expressed in the tall fescue-*E. coenophiala* symbiosis. However, only 1 gene, identified as from a costars family gene (Pang et al., 2010) and corresponding to TFF17 in the previous study (Johnson et al., 2003), was found to have statistically significant differential expression (up-regulated) in S1139RC. The differences between this study and the previous studies may be due to the different species used as well as the different approaches for identification of differentially expressed genes.

Table 2.4. Gene ontology (GO) categorization of the 209 differentially expressed plant genes found by SOLiD-SAGE.

	GO Term	Up-regulated genes		Down-regulated genes	
		Sequences (n)	%	Sequences (n)	%
1	Autophagy	4	6	0	0
2	Carbohydrate metabolic process	0	0	2	1
3	Cell redox homeostasis	0	0	4	3
4	Cellular carbohydrate metabolic process	0	0	1	1
5	Cellular component organization	3	4	3	2
6	Cellular nitrogen compound metabolic process	2	3	2	1
7	DNA modification	2	3	0	0
8	DNA replication	0	0	1	1
9	Generation of precursor metabolites and energy	1	1	5	4
10	Lipid metabolic process	0	0	7	5
11	Meiosis	1	1	0	0
12	Metabolic process	3	4	2	1
13	Organ senescence	2	3	0	0
14	Oxidation reduction process	1	1	1	1
15	Photosynthesis	9	13	1	1
16	Protein metabolic process	2	3	0	0
17	Protein modification process	4	6	6	4
18	Protein targeting	3	4	3	2
19	Protein ubiquitination	1	1	1	1
20	Regulation of hydrolase activity	2	3	0	0
21	Response to abiotic stimulus	1	1	1	1
22	Response to biotic stimulus	1	1	1	1
23	Response to endogenous stimulus	0	0	1	1
24	Response to metal ion	1	1	1	1
25	Response to stress	8	11	10	7
26	RNA metabolic process	0	0	6	4
27	Signal transduction	2	3	8	6
28	Transcription, DNA-dependent	1	1	5	4
29	Translation	0	0	9	7
30	Transport	6	8	19	14
31	Biological process	3	4	19	14
32	No BLASTx match found	7	10	9	7
33	No match in NCBI database	2	3	9	7
	Total	**72**		**137**	

E. festucae gene expression *in planta*

The SOLiD-SAGE tags from S1139RC also included transcripts for endophyte-expressed genes, which were identified by mapping to the fungal reference dataset. Since the whole genome sequence of *E. festucae* was included in the reference dataset, the fungal reference used for mapping could be expected to have near complete coverage of the fungal transcriptome. The fungal biomass in infected plants based on relative amounts of total DNA has been estimated to be less than 2% (Young et al., 2005). The number of fungal mapped tags ranged from 1.4% to 2.8% of the total potential mappable tags, so the percentage of mapped fungal tags is in the expected range. There were 6,298 unique fungal tags that were mapped. The number of genes in the *E. festucae* genome has been estimated to be 9,440 (Schardl et al., 2009) so the SOLiD-SAGE captured tags for approximately 67% of the genes.

The fungal mapped tags were normalized as percent mapped tags in each replicate. From the means of the normalized number of tags we ranked and identified, where possible, the most abundant fungal tags, which were arbitrarily designated as those that were 0.01% and above of the total fungal mapped tags. There were 191 such tags, which accounted for 58% of the total fungal mapped tags. A complete list of the abundant fungal tags can be found as Supporting Information Table S1C in Ambrose and Belanger (2012). The remaining 42% of the fungal mapped tags constituted the bulk of the total fungal genes represented by the SAGE tags, but for many of these there was only a single tag. Identification of the *in planta* abundant fungal transcripts can reveal the metabolic processes to which the fungus is devoting its energy in the symbiotic interaction. The top 20 fungal tags are summarized in Table 2.5.

Table 2.5. The 20 most abundant *E. festucae* transcripts found in the endophyte-infected plant S1139RC.

	Mapped match	Gene identification	Mean % mapped tags
1	SRR493691.19548	NC12 (*Epichloë* specific)	10.12
2	SRR493691.12929	Secreted, Antifungal protein (small, cysteine-rich)	6.34
3	SRR493690.59639	Secreted, Unknown function (small, cysteine-rich)	4.60
4	SRR493691.55163	Secreted, Subtilisin-like protease	2.60
5	SRR493691.16186	Unknown function (*Epichloë* specific)	2.58
6	SRR493691.34496	Secreted, Unknown function (*Epichloë* specific)	1.90
7	SRR493691.15406	Secreted, NC25 (gigA) (*Epichloë* specific)	1.45
8	SRR493691.46417	Secreted, Conidiation associated	1.37
9	SRR493691.19548	Antisense, NC12 (*Epichloë* specific)	1.17
10	SRR493691.31334	Unknown function (*Epichloë* specific)	1.14
11	SRR493691.38494	Unknown function, glucose repressible Grg1	1.04
12	SRR493690.32961	Secreted, Unknown function	1.02
13	SRR493691.32405	Secreted, Unknown function (small, cysteine-rich) (*Epichloë* specific)	1.01
14	SRR493691.6870	Secreted, Unknown function (small, cysteine-rich) (*Epichloë* specific)	0.97
15	AFRX01000547	Secreted, Unknown function (small, cysteine-rich)	0.87
16	SRR493691.52692	Unknown function (*Epichloë* specific)	0.84
17	SRR493691.35651	Secreted, Unknown function (*Epichloë* specific)	0.76
18	SRR493691.71157	Mismatched base pair and cruciform DNA recognition protein	0.62
19	SRR493691.48167	Secreted, Unknown function (small, cysteine-rich) (*Epichloë* specific)	0.62
20	SRR493691.69494	Secreted, Unknown function, RNase domain	0.61

The most abundant endophyte transcript expressed in the endophyte-infected plant was for a protein of unknown function. A similar transcript was reported in the *E. coenophiala*-tall fescue symbiosis and was designated NC12 (Johnson et al., 2003). The corresponding gene appears to be unique to *Epichloë* spp., since similar genes are apparently not present in any other genera for which sequence information is available. This transcript was present at exceptionally high levels in the endophyte-infected strong creeping red fescue, over 10% of the total fungal mapped SAGE tags. In contrast, the most abundant plant tag, the chlorophyll a/b binding protein, was about 1% of the mapped plant tags. The N-terminus of the encoded protein is not yet known, so the

predicted cellular location of the protein cannot be determined.

The second most abundant fungal transcript, 6.34% of the fungal mapped tags, is similar to genes for secreted antifungal proteins from *Penicillium* and *Aspergillus*. Some of these proteins have been characterized and were shown to have antifungal activity against different target organisms as well as having different modes of action (Marx, 2004; Meyer 2008). Some of these antifungal proteins were found to have activity against plant pathogens, and their use in plant protection has been proposed (Marx, 2004; Meyer 2008). Such genes have not previously been recognized in the fungal endophytes of grasses, perhaps because they are not present in most of the *Epichloë* spp. for which genome sequence information is available (Schardl et al., 2013) (discussed more below).

The fourth most abundant tag was for a secreted subtilisin-like protease, which is abundantly expressed in another system, *Epichloë poae* infected *Poa secunda* subsp. *juncifolia* (synonym *Poa ampla* Merr.) (Reddy et al., 1996; Tadych et al., 2012). Another abundant transcript was for a secreted protein that was previously found, along with NC12 described above, in the *E. coenophiala*-tall fescue symbiosis and was designated NC25 (Johnson et al., 2003). NC25, also designated *gigA*, encodes a secreted protein that is cleaved post-translationally to generate multiple cyclic oligopeptides, although the function of the oligopeptides is not yet known (Johnson et al., 2010).

Overall, one striking feature of the abundantly expressed fungal transcripts is that several transcripts are exceptionally abundant and a few fungal transcripts account for a large percentage of the total fungal tags; the top 5 fungal transcripts constituted over 26% of the total mapped tags. Twenty-six of the 191 most abundant fungal tags are for proteins that are apparently *Epichloë* specific. Fifty-nine of the most abundant tags are for secreted proteins, most of which are of

unknown function. Nineteen of the 59 secreted proteins, including the antifungal protein discussed above, can be characterized as small (less than 210 amino acids) cysteine-rich proteins. Small secreted cysteine-rich proteins are involved in the interactions of fungal pathogens with their hosts, including some Avr proteins (Templeton et al., 1994; Rep, 2005; Rouxel et al., 2012), and may also have a role in the endophyte-host symbiosis.

One of the secreted small cysteine-rich proteins (0.1%) had high similarity to the virally-encoded antifungal killer protein 4 (KP4) secreted by *Ustilago maydis*. KP4 shows antifungal activity by primarily disrupting target calcium uptake to stunt growth. Inter-kingdom horizontal gene transfer has been suggested for the presence of the protein in virus, moss and fungi (Brown, 2011). KP4, along with the antifungal protein discussed above, may be a candidate gene for *Epichloë*-mediated disease resistance, although KP4-like genes are present in all *Epichloë* genomes sequenced thus far (Schardl et al., 2013).

Endophyte transcripts putatively involved in plant defense evasion

Other fungal genes that may play a role in the symbiotic association with the host were also among the highly expressed genes. A gene annotated as a salicylate hydroxylase was 0.02% of the mapped tags. Salicylate hydroxylase converts salicylate, which is required for induction of systemic acquired resistance in plants, to catechol, which is not effective (Gaffney et al., 1993). Expression of an *E. festucae* salicylate hydroxylase gene could be a mechanism of suppressing the host's defense response. A *Fusarium* sp. was identified that could use salicylate as its sole carbon source and salicylate hydroxylase activity was detected (Dodge and Wackett, 2005). A detailed study of the *E. festucae* salicylate hydroxylase gene identified here is described in Chapter 4 of this thesis.

Two SAGE tags for secreted LysM domain proteins constituted 0.11% and 0.03% of the fungal mapped tags. Secreted LysM domain effector proteins from the plant pathogens *Cladosporium fulvum*, *Mycosphaerella graminicola*, and *Magnaporthe oryzae* have been shown to bind chitin and thereby suppress the chitin-triggered plant defense responses (de Jonge et al., 2010; Marshall et al., 2011; Mentlak et al., 2012). Such proteins may have a similar role in plant-fungal symbiotic systems. A transcript for a LysM domain protein was among the most highly expressed transcripts in the symbiotic interaction of the mycorrhizal fungus *Tuber melanosporum* with the host roots and was proposed to play a role in sequestering chitin molecules from the host defense system (Plett and Martin, 2012).

SAGE tags for a major facilitator superfamily (MFS) transport protein were 0.03% of the fungal mapped tags. Comparison of the corresponding protein sequence with a transporter database (http://www.membranetransport.org/) (Ren et al., 2007) identified the *E. festucae* Rose City protein as similar to tetracycline and multidrug efflux transporters. Such efflux transporters could be for transport of fungal compounds, perhaps an alkaloid, or toxic plant defense compounds out of the cell.

Nutrient acquisition mechanisms *in planta*

How the endophyte obtains its nutrients from within the host plant apoplast is not yet understood. Tags for transporters for nitrogen and carbon compounds were among the abundant tags, suggesting the transported substrates are important nutrient sources for the endophyte. Tags for an ammonium permease (0.02%) similar to the high affinity MepA permease of *Fusarium fujikuroi* (Teichert et al., 2008) and 4 amino acid permeases were among the abundant SAGE tags. Ammonium and amino acids are expected to be present in the apoplast (Kuldau and Bacon, 2008), so both sources of nitrogen are apparently being utilized by *E. festucae*. The *E. festucae* genome (Schardl et al., 2013) contains 21 genes annotated as amino acid permeases, but SAGE tags were

recovered for only 9; 5 of which were of low abundance (<0.005%). Based on comparison with the functionally characterized amino acid permeases of the *Saccharomyces cerevisiae* (Regenberg et al., 1999) the abundantly expressed *E. festucae* genes were annotated as a proline permease similar to the proline specific Put4p permease (0.03%), 2 different arginine permeases (0.02% each), and an amino acid permease related to a choline transporter (0.01%).

Tags for four transporter proteins that may play a role in fungal carbon acquisition were also among the most abundant fungal tags. A malic acid/C4-dicarboxylate transporter was 0.08% of the mapped tags. The *E. festucae* Rose City malic acid/C4-dicarboxylate transporter is similar to that of *Schizosaccharomyces pombe*, which was functionally characterized and shown to be involved in uptake of malate, succinate and malonic acid (Grobler et al., 1995). Malate is known to be present in the apoplast (Kuldau and Bacon, 2008; Fernie and Martinoia, 2009; Meyer et al., 2010), so the abundance of SAGE tags for a malate transporter suggests that malate may serve as a carbon source for the fungus. Malate levels were increased in endophyte-infected ryegrass and malate was proposed to be important for lipid biosynthesis in the lipid storing *Epichloë* (Rasmussen et al., 2008).

The predominant sugars expected to be present in the apoplast are sucrose, glucose, and fructose (Kuldau and Bacon, 2008). SAGE tags for a gene similar to a functionally characterized fructose specific transporter from *Botrytis cinerea* (Doehlemann et al., 2005) were 0.02% of the mapped tags. A glucose/xylose transporter was also 0.02% of the fungal mapped tags and is similar to transporters characterized from *Candida intermedia* (GXS1) and *Colletotrichum graminicola* (CgHXT1) (Leandro et al., 2006; Lingner et al., 2011). The *C. intermedia* protein was shown to transport both glucose and xylose and the *C. graminicola* protein transported glucose, mannose, fructose, and xylose. The possibility that the *E. festucae* protein could transport xylose in addition to glucose is of interest since xylose is abundant in the hemicellulose component of grass cell

walls (Scheller and Ulvskov, 2010). A similar, yet distinct (since the best match to the *E. festucae* genome is on a different contig), transporter from *E. lolii* was functionally characterized and shown to preferentially transport mannose into the cell (Rasmussen et al., 2012). The authors also detected the presence of a fungal mannosidase in the infected plant and therefore hypothesized the endophyte has the capability to access cell wall carbohydrates. SAGE tags for an *E. festucae* transcript similar to the characterized mannose transporter were present, but at lower abundance (0.002%) than those for the glucose/xylose transporter (0.02%).

The possibility of *E. festucae* using xylose as a carbon source *in planta* is supported by the abundance of SAGE tags for enzymes required for utilization of xylose. A secreted ß-xylosidase (0.01%) could hydrolyse extracellular xylo-oligosaccharides to xylose (de Vries and Visser, 2001), which could then be transported into the cell via the glucose/xylose transporter (0.02%). The cytoplasmic enzymes xylose reductase (0.02%), xylitol dehydrogenase (0.01%), and xylulose kinase (0.01%) together could act to convert xylose to xylulose-5-phosphate, which could then enter the pentose phosphate pathway (Bettiga et al., 2008). *E. festucae* was able to grow in culture with xylose as the sole carbon source (Li et al., 2008).

SAGE tags for a gene similar to the functionally characterized *Saccharomyces cerevisiae* alpha-glucoside transporter AGT1 were 0.01% of mapped tags. The yeast AGT1 transporter has high affinity for sucrose and lower affinity for maltose and maltotriose (Batista et al., 2004; Alves Jr et al., 2008). Sucrose transported into the fungal cell could then be broken down to glucose and fructose by a cytoplasmic invertase (0.01%). SAGE tags for a secreted invertase were recovered at a lower level (0.002%). Apparently *in planta* the endophyte can gain access to the host synthesized sucrose both through direct uptake followed by cytoplasmic hydrolysis and through apoplastic hydrolysis, followed by uptake of the derived glucose and fructose, as was previously proposed for *E. festucae* in culture (Lam et al., 1994).

Reactive oxygen species in the fungal-plant symbiosis

The generation of and protection from reactive oxygen species (ROS) are critical factors in plant-pathogen interactions (Torres, 2010; Heller and Tudzynski, 2011) and are also critical in the endophyte-grass symbiosis (Scott and Eaton, 2008; White and Torres, 2010; Hamilton et al., 2012; Tanaka et al., 2012). Production of reactive oxygen species by *E. festucae* through the activity of the NADPH oxidase NoxA, is known to be an important factor in the maintenance of the mutualistic symbiosis, probably by regulating hyphal growth and branching of the fungus *in planta* (Tanaka et al., 2006; Tanaka et al., 2008; Scott and Eaton, 2008). SAGE tags for the *E. festucae* NoxA were recovered at 0.003%.

SAGE tags for other genes relevant to ROS were among the abundant tags. From a proteomic analysis high levels of a fungal cytoplasmic superoxide dismutase were found in the *E. lolii-Lolium perenne* symbiosis (Zhang et al., 2011). The authors proposed the enzyme was important in protecting the endophyte from reactive oxygen species. SAGE tags for a similar superoxide dismutase (0.11%) as well as a copper chaperone for superoxide dismutase (0.01%) (Culotta et al., 1997) were among the abundant tags, consistent with the previous study. SAGE tags for other genes involved in protection from oxidative stress were glutaredoxin (0.15%), glutathione S-transferase (0.05%), and a secreted thioredoxin reductase (0.02%). SAGE tags for a secreted galactose oxidase, which generates hydrogen peroxide, were 0.08%. Galactose is typically present in the apoplast (Kuldau and Bacon, 2008) so galactose oxidase may be a source of the endophyte ROS proposed to enhance leakage of nutrients from the host cells and to induce plant synthesis of antioxidants (White and Torres, 2010).

Absence of highly expressed transcripts for fungal alkaloids

One of the major benefits of endophyte infection to the grass host is protection from insect and mammalian herbivory mediated by the synthesis of toxic alkaloids, which are generally found *in*

planta and not in culture. There are four classes of alkaloids associated with protection from herbivory and *E. festucae* can produce all four, although no single isolate is known that produces all four (Schardl, 2001). The endophyte-infected plant genotype used in this study, S1139RC, was previously analyzed for three of the alkaloid classes. The ergot alkaloid ergovaline and the indole-diterpene lolitrem B were detected, but there was no peramine detected (Yue et al., 2000). The pyrrolopyrazine alkaloid peramine, an insect feeding deterrent, is synthesized via a non-ribosomal peptide synthetase (*perA*) (Tanaka et al., 2005).

There was an *E. festucae* 454 sequence that was a good match to the *perA* gene, however, there was no SAGE tag corresponding to the *perA* gene. Whether the Rose City isolate of *E. festucae* can actually produce peramine is not yet known. The indole-diterpenes and the ergot alkaloid ergovaline are mammalian mycotoxins that each require 11 genes for their biosynthesis (Schardl et al., 2012,Young et al., 2005; Yong et al., 2006; Fleetwood et al., 2007). The loline alkaloids have insecticidal properties and also require 11 genes for their synthesis (Schardl et al., 2012; Kutil et al., 2007). None of the genes for these 3 classes of alkaloids were represented in either of the 454 sequence datasets or the SAGE tags, although this *E. festucae* isolate is known to produce both lolitrem and ergovaline. The expression level of the alkaloid biosynthetic genes in S1139RC must be relatively low, such that their transcripts were not captured in the sequencing described in this study.

The highly expressed *E. festucae* antifungal protein gene is not present in many *Epichloë* spp.

The second most abundantly expressed fungal protein stood out as being of interest regarding the disease resistance conferred on the host plant by *E. festucae*. This protein has been designated as an antifungal protein and was similar to proteins from *Penicillium* and *Aspergillus*, that have been demonstrated to have antifungal properties (Marx, 2004). Since the *Epichloë* endophyte-mediated

disease resistance is unique to the fine fescues infected with *E. festucae*, presumably there is some unique feature of those symbioses that is responsible for the disease resistance.

Whole genome sequences are available for 10 *Epichloë* species: *E. festucae* E2368 (Accession ADFL02000000), *E. festucae* Fl1 (Accession AFRX01000000), *E. amarillans* E57 (Accession AFRF01000000), *E. brachyelytri* E4804 (Accession AFRB01000000), *E. glyceriae* E277 (Accession AFRG01000000), *E. typhina* E5819 (Accession AFSE01000000), *E. gansuensis* E7080 (Accession AFRE01000000). Sequences for *E. elymi*, *E. typhina* E8 and *E. gansuensis* var. *inebrians* are available at http://csbio-l.csr.uky.edu (Schardl et al., 2013). One of the *E. festucae* Rose City 454 sequences with the complete coding sequence of the antifungal protein gene (Accession SRR493691.12929) was used in BLASTn and tBLASTn searches of the *Epichloë* genomes. Matches were found only in *E. festucae* E2368 and *E. gansuensis* var. *inebrians*, suggesting the other *Epichloë* genomes do not have a gene for the antifungal protein. Nothing is known regarding whether infection by *E. gansuensis* var. *inebrians* confers disease resistance to its host grass, *Achnatherum inebrians*. Whether the *E. festucae* antifungal protein is a factor in the reported disease resistance of endophyte-infected *F. rubra* will require additional studies, but based on its high level of expression it should be considered a candidate gene.

Nonetheless, the presence of an abundantly expressed gene, similar to antifungal protein genes from *Penicillium* and *Aspergillus,* apparently only in 2 of the *Epichloë* genomes is interesting and raises the question of whether the gene was the result of gene gain or gene loss. Gene gain or loss can often be inferred by placing presence or absence of the gene on a species phylogeny, as well as by comparing a species phylogeny with a gene phylogeny (Keeling and Palmer, 2008). A species phylogeny based on the conserved MCM7 gene, a subunit of the hexomeric minichromosome maintenance complex (MCM), which is involved in DNA replication initiation (Alabert and Groth,

2012) was reconstructed. MCM7 was previously shown to generate a robust fungal phylogeny across a wide evolutionary distance (Aguileta et al., 2008).

Maximum parsimony phylogenetic analysis of MCM7 sequences from fungal species within the genera that have antifungal protein genes, as well as some related species, is shown in Figure 2.2A. The sources of the sequences used in the analysis are listed in Table 2.5. The tree was based upon 2,480 total characters, of which 870 were constant, 142 variable characters were parsimony uninformative, and 1,468 variable characters were parsimony informative. The sequence from the Pezizomycete *Tuber melanosporum* was designated the outgroup for rooting the tree. The relationships of the MCM7 sequences from the selected fungal species are as expected based on fungal species phylogeny (Hibbett et al., 2007). The presence or absence of an antifungal protein gene similar to that found in *E. festucae* in each of the fungal species included in the MCM7 phylogenetic tree is indicated. The distribution of the antifungal protein gene was patchy, with some genera having species that both have and do not have the gene.

Of the 10 *Epichloë* for which whole genome sequences are available only one other, that of *E. gansuensis* var *inebrians*, had an antifungal protein gene. Similarly to the situation with the two sequenced *E. festucae* isolates, *E. gansuensis* var. *inebrians* has an antifungal protein gene whereas the closely related but morphological distinct isolate *E. gansuensis* (Moon et al., 2007) did not.

A maximum parsimony phylogenetic analysis of the mRNA sequences of the antifungal proteins revealed some unexpected relationships (Figure 2.2B). The annotation of the *E. festucae* antifungal protein gene was modified from that presented with the genome sequence and is based on the 454 sequence from this study. Extensive searches of the NCBI databases have found similar genes in only two classes of fungi, the Eurotiomycetes and Sordariomycetes, with the exception of

Pyrenophora spp., which are in the class Dothideomycetes. In order to carry out an exhaustive search, sequences from some of the species were excluded from the analysis if there was a similar sequence from a closely related species. There is a similar sequence in the *Claviceps paspali* genome that was not included in the analysis because it may be a pseudogene since annotation would require imposing a nonconsensus intron splice site in order to generate the predicted protein sequence.

Since there was no appropriate sequence with which to root the tree, it was midpoint rooted. The tree from the exhaustive search was based on 297 total characters of which 62 were constant, 44 variable characters were parsimony uninformative, and 191 variable characters were parsimony informative. The antifungal protein sequences that grouped in clades 1b and 2 conform to the species phylogeny in that the Eurotiomycetes species and the Sordariomycetes species are in separate clades. However, clade 1a, which includes the *E. festucae* sequence, deviates from the species phylogeny in that it includes a *Pennicillium chrysogenum* antifungal protein sequence. *P. chrysogenum* was the only species identified that had two antifungal protein genes in its genome. Clearly, the antifungal protein gene has a complex evolutionary history. The patchy distribution of the gene among species within the same genus and the discordance of clade 1a in the gene phylogeny with the species phylogeny are suggestive of numerous instances of gene loss as well as the possibility that the *P. chrysogenum* gene in clade 1a was the result of horizontal gene transfer.

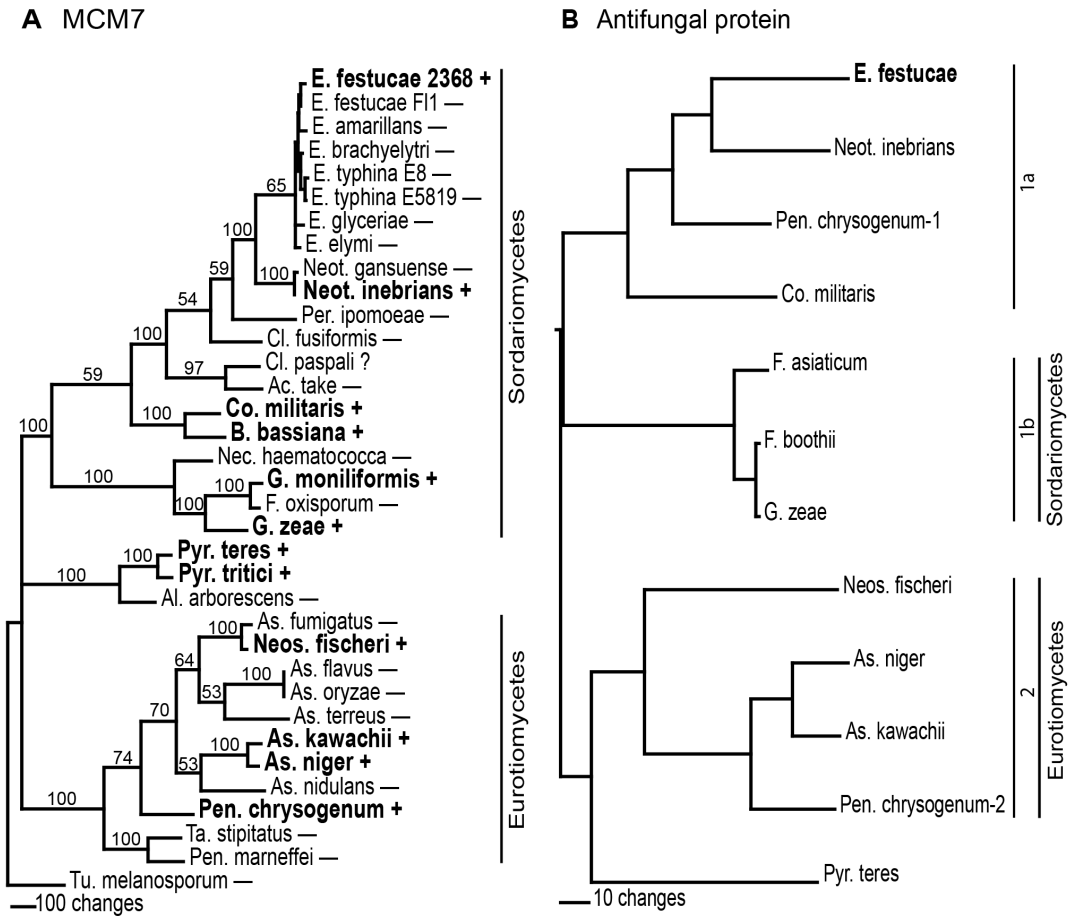


Figure 2.2. The phylogenetic relationships of the MCM7 and antifungal protein coding sequences. (**A**) Rooted 50% majority rule maximum parsimony phylogenetic tree of the MCM7 coding sequences. The *Tu. melanosporum* sequence was designated as the outgroup for rooting the tree. The numbers at the nodes are the bootstrap percentages based on 1,000 replications. The presence (+) or absence (-) of an antifungal protein gene is indicated for each species in the tree. (**B**) The single most parsimonious phylogenetic tree recovered from an exhaustive search of the antifungal protein coding sequences. The tree is midpoint rooted. Accession numbers of the sequences used for both trees are given in Table 2.5.

Table 2.6. Accession numbers of the MCM7 and antifungal protein sequences used in the phylogenetic analyses presented in Figure 2.2. NA indicates not applicable.

Species	Accession number	
	MCM7	**Antifungal protein**
Aciculosporium take[1]	AFQZ01000467	NA
Alternaria arborescens[1]	AIIC01000090	NA
Aspergillus flavus[1]	XM_002377869	NA
Aspergillus fumigatus[1]	XM_750254	NA
Aspergillus kawachii[1]	BACL01000094	GAA86026
Aspergillus nidulans[1]	XM_658504	NA
Aspergillus niger[1]	XM_001397760	XP_001391221
Aspergillus oryzae[1]	XM_001826176	NA
Aspergillus terreus[1]	XM_001213626	NA
Beauveria bassiana[1]	ADAH01000069	ADAH01000656
Claviceps fusiformis[1]	AFRA01000297	NA
Claviceps paspali[1]	AFRC01000014	NA
Cordyceps militaris[1]	AEVU01000284	EGX93824
Epichloë amarillans[1]	AFRF01000144	NA
Epichloë brachyelytri[1]	AFRB01001052	NA
Epichloë elymi[1]	Schardl et al. (2013)	NA
Epichloë festucae 2368[1]	ADFL02000041	SRR493691.12929
Epichloë festucae Fl1[1]	AFRX01000012	NA
Epichloë gansuense[1]	AFRE01000016	NA
Epichloë gansuense var *inebrians*[1]	Schardl et al. (2013)	Schardl et al. (2013)
Epichloë glyceriae[1]	AFRG01000147	NA
Epichloë typhina E5819[1]	AFSE01000068	NA
Epichloë typhina E8[1]	Schardl et al. (2013)	NA
Fusarium asiaticum	NA	CAR79023
Fusarium avenaceum	NA	CAR79018
Fusarium boothii	NA	CAR79010
Fusarium cerealis	NA	CAR79014
Fusarium oxysporum[1]	AGNB01000011	NA
Fusarium poae	NA	CAR79017
Gibberella moniliformis[1]	AAIM02000101	AAIM02007600
Gibberella zeae[1]	AACM02000242	XP_384921
Nectria haematococca[1]	XM_003051643	NA
Neosartorya fischeri[1]	XM_001260497	XP_00126258
Penicillium chrysogenum[1]	XM_002565874	XP_002557660
Penicillium marneffei[1]	XM_002146315	NA
Periglandula ipomoeae[1]	AFRD01000429	NA
Pyrenophora teres[1]	AEEY01004277	AEEY01001915
Pyrenophora tritici-repentis[1]	AAX101000415	XM_001934290
Talaromyces stipitatus[1]	XM_002478600	NA
Tuber melanosporum[1]	CABJ01000538	NA

[1]Species for which whole genome sequences are available (Schardl et al., 2013).

Expression of an alternatively spliced variant of the *E. festucae* antifungal protein gene

We also detected a SAGE tag that mapped to an annotated alternatively spliced variant of the *E. festucae* antifungal protein (*Ef*-AFP) in the *E. festucae* E2368 genome (Schardl et al., 2013). The splice variant SAGE tag is 0.03% of the total *E. festucae* mapped tags while the *Ef*-AFP tag is 6.34% of the mapped tags. The splice variant coding sequence extends downstream of the more abundant *Ef*-AFP (Figure 2.3A). The presence of a transcript corresponding to the SAGE tag of the alternatively spliced variant was confirmed by RT-PCR of cDNA from SR1139RC (Figure 2.3B). The genome annotation predicts an intron within the amplified region. The sequence of the RT-PCR product was identical to the genome sequence, indicating there is no intron in this region. Extensive genome searches of the NCBI databases did not find any other organisms with a similar potential alternatively spliced variant in their genome.

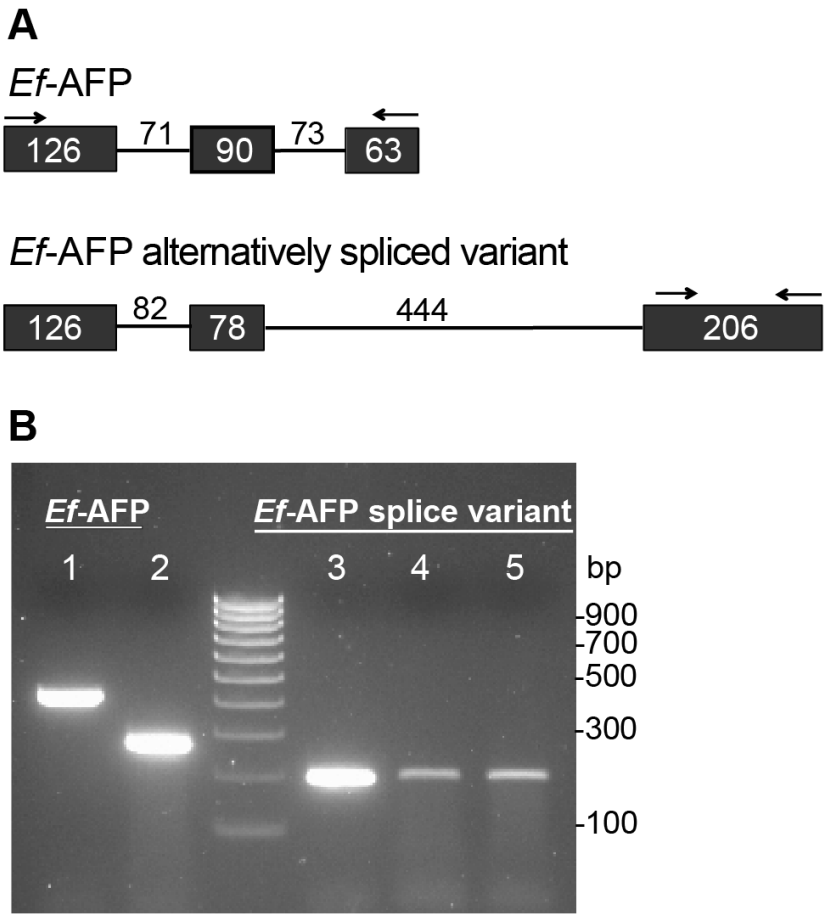


Figure 2.3. Detection of an alternatively spliced variant of *Ef-AFP*. (**A**) Structural features of the *Ef-AFP* gene and the alternatively spliced variant. Exons are depicted as dark boxes, introns as lines, and sizes are given in bp. Arrows indicate positions of primers used for PCR amplification. (**B**) PCR products of (1) *Ef-AFP* using *E. festucae* genomic DNA, and (2) *E. festucae*-infected plant cDNA generated from oligo(dT) as templates; (3) partial *Ef*-AFP alternatively spliced variant using *E. festucae* genomic DNA, (4) *E. festucae*-infected plant cDNA generated from a gene-specific primer (Alt *Ef-AFP* Reverse), and (5) *E. festucae*-infected plant cDNA generated from oligo(dT) as templates.

Antisense SAGE tags

Antisense transcripts have been found in many organisms and are considered to have regulatory roles in gene expression (Faghihi and Wahlestedt, 2009; Werner and Swan, 2010). However, the exact mechanism of action and function of these molecules remains unclear. The directionality of SAGE tags allows the detection of antisense transcripts. Antisense SAGE tags have been reported in rice, *A. thaliana*, wheat, sugarcane, and human transcriptome studies (Gibbings et al., 2003; Robinson et al., 2004; Quéré et al., 2004; Ge et al., 2006; Calso and Figueira, 2007; Poole et al., 2008).

Eight antisense tags were among the 191 most abundant fungal tags. The most abundant antisense tag (1.17% of mapped tags) was for NC12, which also was the most abundant sense tag (10.12%). Other abundant fungal antisense tags were for NC25 (0.13%), plasma membrane proteolipid (0.04%), subtilisin-like protease (0.03%), and four genes of unknown function (0.18%, 0.12%, 0.08%, 0.05%). The abundance of the fungal antisense tags relative to the corresponding sense tag ranged from 1:9 (NC12) to 1:87 (subtilisin-like protease). One of the antisense tags for a gene of unknown function (0.05%) did not have a matching sense tag.

Since fungal antisense tags were identified, the plant SAGE tags were searched for the presence of antisense tags. Antisense tags were identified among the plant SAGE tags for metallothionein, chlorophyll a-b binding protein, and non-specific lipid-transfer protein. The plant antisense tags were present at low abundance, ranging from 16 (non-specific lipid transfer protein) to 85 (chlorophyll a-b binding protein) per MMT. A plant antisense tag for omega 6 fatty acid desaturase was among the differentially expressed tags, as was the sense tag for the same transcript. The abundance of the plant antisense tags relative to the corresponding sense tags ranged from 1:140 (chlorophyll a-b binding protein) to 1:373 (non-specific lipid-transfer protein), considerably lower than that of the fungal antisense tags discussed above. Reverse transcriptase

PCR (RT-PCR) using strand-specific primers for cDNA synthesis was used to confirm the presence of antisense plant and fungal transcripts in RNA isolated from S1139RC (Figure 2.4).

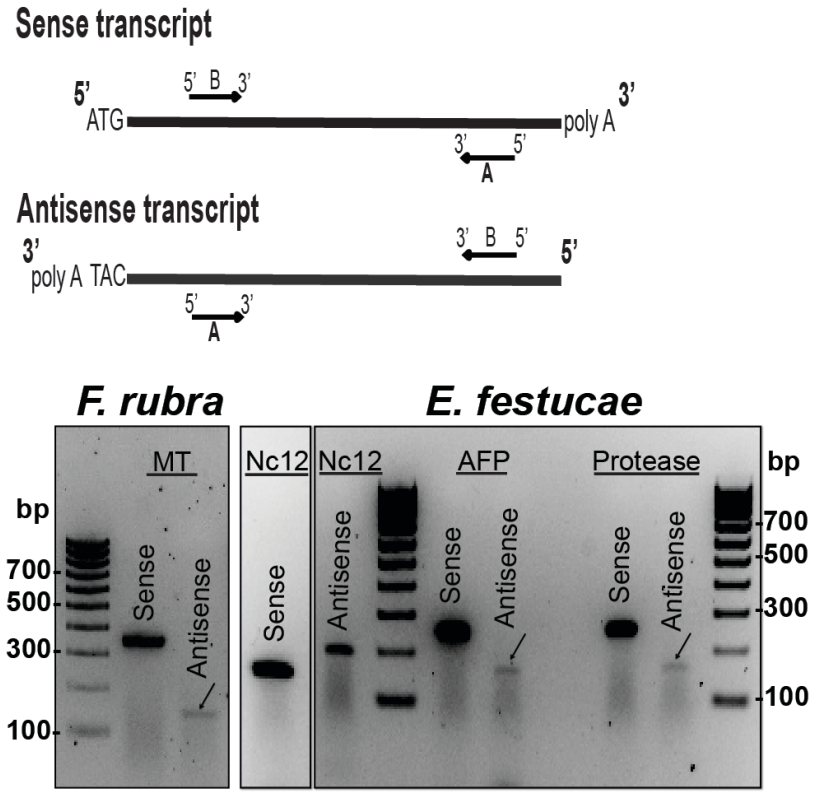


Figure 2.4. Gel analysis of *F. rubra* and *E. festucae* antisense transcripts. The diagram illustrates primer design for detection of sense and antisense transcripts. The "A" primers were used for strand specific synthesis of cDNA from the RNA sample. The "A" and "B" primers were used for cDNA amplification. cDNAs generated from gene-specific primers for the *F. rubra* metallothionein (MT) and the *E. festucae* NC12, antifungal protein (AFP), and subtilisin-like protease were used as templates for PCR amplification.

Discussion

Since nothing is yet known regarding the biological basis of the endophyte-mediated disease resistance in *F. rubra*, possibilities are that it may be due to fungal genes and/or fungal induction of plant genes. As a first step in determining the effect of the presence of *E. festucae* on host gene expression, a quantitative transcriptome comparison between endophyte-free and endophyte-infected plants using SOLiD-SAGE was done. The results obtained in this study highlight the usefulness of next generation sequencing technologies for non-model organisms, such as the *Festuca rubra–Epichloë festucae* symbiotic system. There is no genome sequence for *F. rubra* and, prior to this study, only a modest number (1,773) of EST sequences was available. For relatively low cost greater than 100,000 454 EST sequences were generated. The ESTs were conservatively estimated to represent 56% of the *F. rubra* genes. Having such a homologous reference dataset greatly facilitated the ability to map the SOLiD-SAGE tags. With a more complete plant gene reference dataset, it is likely that more SAGE tags would have been mapped. In a similar study with the non-model organism *Nanomia bijuga* (Cnidaria) in which an incomplete 454 reference dataset was used, 27% of the SAGE tags were mapped (Siebert et al., 2011).

In many next-generation transcriptome studies aimed at evaluating differential expression between samples, the importance of biological replication has not been recognized (Auer et al., 2011). Statistical analysis on deep sequencing of single replicates is not a replacement for biological replication in a differential expression study. Here, three biological replicates of the E- and E+ samples were used in order to identify those SOLiD-SAGE tags whose representation in the two samples was statistically significantly different. Even with an incomplete reference dataset, over two hundred plant genes were identified that showed statistically significant differential gene expression due to the presence of the fungal endophyte. Since numerous plant genes were differentially expressed, clearly the host is sensing the presence of the endophyte in some way.

The affected plant genes are involved in a wide variety of physiological processes. These results indicate that the presence of the endophyte has moderate effects on plant metabolism in many different processes. However, understanding how the observed effect on expression of individual plant genes is correlated with the overall benefits of endophyte-infection to the host will be challenging.

The effects of fungal endophyte infection on host grass metabolism are complex and influenced by the specific endophyte-host genotype interaction as well as by environmental conditions (Rasmussen et al., 2008). The plants used in this study were grown under growth chamber conditions in the absence of any biotic or abiotic stress. Analysis of the SAGE tags revealed hundreds of plant genes whose expression levels were moderately affected by the presence of the fungal endophyte, although there was no phenotypic difference between the endophyte-free and endophyte-infected plants. Dramatic differences in plant phenotype are often seen in field studies comparing endophyte-free and endophyte-infected plants experiencing stress, such as drought, poor soil conditions, or disease (Kuldau and Bacon, 2008; Clarke et al., 2006). Future transcriptome comparisons of endophyte-free and endophyte-infected plants experiencing stress may reveal larger differences in plant genes affected by the presence of the endophyte than those reported here. The results reported here on the fine fescue-fungal endophyte symbiosis is a first step in ultimately understanding how the differential expression of individual plant genes in response to endophyte infection results in changes in host metabolism.

The quantitative analysis of the fungal gene expression reported here revealed new information on the endophyte transcriptome *in planta*. A striking feature of the fungal SOLiD-SAGE tags was that some transcripts were present at surprisingly high levels and that many of the abundant transcripts were for secreted proteins. Fifty-nine of the 191 most abundant fungal transcripts, representing almost 31% of the abundantly expressed genes, were for secreted proteins. The size of the

predicted secretomes of 9 fungal species was found to be between 5-12% of the total proteome (Soanes et al., 2008), so the *E. festucae* secreted proteins are highly overrepresented in the abundant SAGE tags. The functions of most of these secreted proteins are unknown and some appear to be unique to the *Epichloë*. Since these abundant secreted proteins are present at the interface of the two organisms, it is likely they are important in the symbiotic interaction between plant and fungus. Many of the unknown secreted proteins could be characterized as small, secreted, cysteine-rich proteins, which are known to be important in plant pathogen interactions (Rep, 2005) as well as in cell-to-cell signaling in plant development (Marshall et al., 2011). Also among the most abundant fungal tags were tags for antisense transcripts, which may play a role in regulating the expression of the even more abundant sense transcripts.

Two outstanding questions in understanding the *Epichloë*-grass symbiotic system are, how does the fungus evade the plant defenses, and what is the mechanism of nutrient acquisition by the fungus. Fungal tags for two secreted LysM domain proteins, a salicylate hydroxylase, and an MFS drug efflux transporter, an ammonium permease, 4 amino acid permeases, a malate transporter, a fructose specific transporter, a glucose/xylose transporter, a sucrose transporter and an alpha-glucoside transporter were among the most highly expressed tags and are candidates for involvement in these two critical processes. Sucrose, glucose, and fructose are known to be components of the apoplast (Kuldau and Bacon, 2008) and thus likely carbon sources for the endophyte.

Analysis of the plant genes whose transcript levels were affected by the presence of the fungal endophyte did not reveal any strong candidates for genes directly related to enhanced disease resistance. However, one of the fungal abundant secreted proteins is of particular interest regarding the disease resistance observed in endophyte-infected fine fescues. This protein is similar to characterized antifungal proteins from *Penicillium* and *Aspergillus* (Marx, 2004). The

uniqueness of this gene in *E. festucae* from *F. rubra*, its transcript abundance, and the secreted nature of the protein, all suggest it may be involved in the disease resistance conferred to the host, which is a unique feature of the fine fescue–endophyte symbiosis. The possibility that it may be involved in conferring disease resistance to the host makes this an extremely interesting protein for further research. Overall, this study has identified numerous plant and fungal genes likely to be important in the endophyte-grass interaction that are candidates for future functional characterization.

References

1. Aguileta G, Marthey S, Chiapello H, Lebrun M-H, Rodolphe F, Fournier E, Gendrault-Jacquemard A, Giraud T (2008) Assessing the performance of single-copy genes for recovering robust phylogenies. Systematic Biology, 57:613-627.

2. Alabert C, Groth A (2012) Chromatin replication and epigenome maintenance. Nature Review Molecular Cell Biology, 13:153-167.

3. Alves Jr SL, Herberts RA, Hollatz C, Trichez D, Miletti LC, De Araujo PS, Stambuk BU (2008) Molecular analysis of maltotriose active transport and fermentation by *Saccharomyces cerevisiae* reveals a determinant role for the *AGT1* permease. Applied Environmental Microbiology, 74:1494-1501.

4. Auer PL, Srivastava S, Doerge RW (2011) Differential expression – the next generation and beyond. Briefings in Functional Genomics, doi:10.1093/bfgp/elr041.

5. Batista A, Miletti LC, Stambuk BU (2004) Sucrose fermentation by *Saccharomyces cerevisiae* lacking hexose transport. Journal of Molecular Microbiology and Biotechnology, 8:26-33.

6. Bausher M, Shatters R, Chaparro J, Dang P, Hunter W, Niedz R (2003) An expressed sequence tag (EST) set from *Citrus sinensis* L. Osbeck whole seedlings and the implications of further perennial source investigations. Plant Science, 165:415-422.

7. Bettiga M, Hahn-Hagerdal B, Gorwa-Grauslund MF (2008) Comparing the xylose reductase/xylitol dehydrogenase and xylose isomerase pathways in arabinose and xylose fermenting *Saccharomyces cerevisiae* strains. Biotechnology for, Biofuels 1:16

8. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R (2009) QuickGO: a web-based tool for Gene Ontology searching. Bioinformatics, 25:3045-3046.

9. Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J (2010) Galaxy: a web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology, Chapter 19:Unit 19.10.1-21.

10. Bonos SA, Wilson MM, Meyer WA, Funk CR (2005) Suppression of red thread in fine fescues through endophyte-mediated resistance. Applied Turfgrass Science, 10:1094.

11. Brown DW (2011) The KP4 killer protein gene family. Current Genetics, 57:51-62.

12. Calso T, Figueira A (2007) Serial analysis of gene expression in sugarcane (*Saccharum* spp.) leaves revealed alternative C4 metabolism and putative antisense transcripts. Plant Molecular Biology, 63:745-762.

13. Christensen MJ, Bennett RJ, Ansari HA, Koga H, Johnson RD, Bryan GT, Simpson WR, Koolaard JP, Nickless EM, Voisey CR (2008) *Epichloë* endophytes grow by intercalary hyphal extension in elongating grass leaves. Fungal Genetics and Biology, 45:84-93.

14. Clarke BB, White JF Jr, Hurley RH, Torres MS, Sun S, Huff DR (2006) Endophyte-mediated suppression of dollar spot disease in fine fescues. Plant Disease, 90:994-998.

15. Clay K, Schardl C (2002) Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. The American Naturalist, 160:99-127.

16. Conesa A, Götz S, Garcia-Gomez JM, Terol J, Talon M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics, 21:3674-3676.

17. Conesa A, Götz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. International Journal of Plant Genomics, 1-13.

18. Culotta VC, Klomp LWJ, Strain J, Casareno RLB, Krems B, Gitlin JD (1997) The copper chaperone for superoxide dismutase. Journal of Biological Chemistry, 272: 23469-23472.

19. de Jonge R, van Esse HP, Kombrink A, Shinya T, Desaki Y, Bours R, van der Krol S, Shibuya N, Joosten MHAJ, Thomma BPHJ (2010) Conserved fungal LysM effector Ecp6 prevents chitin-triggered immunity in plants. Science, 329:953-955.

20. de Vries RP, Visser J (2001) *Aspergillus* enzmes involved in degradation of plant cell wall polysaccharides. Microbiology and Molecular Biology Reviews, 65:497-522.

21. Dinkins RC, Barnes A, Waters W (2010) Microarray analysis of endophyte-infected and endophyte-free tall fescue. Journal of Plant Physiology, 167:1197-1203.

22. Doehlemann G, Molitor F, Hahn M (2005) Molecular and functional characterization of a fructose specific transporter from the gray mold fungus *Botrytis cinerea*. Fungal Genetics and Biology, 42:601-610.

23. Dodge AG, Wackett LP (2005) Metabolism of bismuth subsalicylate and intracellular accumulation of bismuth by *Fusarium* sp. strain BI. Appled Environmental Microbiology, 71:876-882.

24. Eaton CJ, Cox MP, Ambrose B, Becker M, Hesse U, Schardl CL, Scott B (2010) Disruption of signaling in a fungal-grass symbiosis leads to pathogenesis. Plant Physiology, 153:1780-1794.

25. Emanuelsson O, Brunak S, von Heijne G, Nielsen H (2007) Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols, 2: 953-971.

26. Faghihi MA, Wahlestedt C (2009) Regulatory roles of natural antisense transcripts. Nature Review Molecular Cell Biology, 10: 637-643.

27. Fernie AR, Martinoia E (2009) Malate. Jack of all trades or master of a few? Phytochemistry, 70:828-832.

28. Fleetwood DJ, Scott B, Lane GA, Tanaka A, Johnson RD (2007) A complex ergovaline gene cluster in *Epichloë* endophytes of grasses. Applied Environmental Microbiology, 73:2571-2579.

29. Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G, Uknes S, Ward E, Kessmann H, Ryals J (1993) Requirement of salicylic acid for the induction of systemic acquired

resistance. Science, 261:754-756.

30. Ge X, We Q, Jung Y-C, Chen J, Wang SM (2006) A large quantity of novel human antisense transcripts detected by LongSAGE. Bioinformatics, 22:2475-2479.

31. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A (2005) Galaxy: a platform for interactive large-scale genome analysis. Genome Research, 15:1451-1455.

32. Gibbings JG, Cook BP, Dufault MR, Madden SL, Khuri S, Turnbull CJ, Dunwell JM (2003) Global transcript analysis of rice leaf and seed using SAGE technology. Plant Biotechnology Journal, 1:271-285.

33. Goecks J, Nekrutenko A, Taylor J, The Galaxy Team (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biology, 11:86.

34. Götz S, Garcia-Gomez JM, Terol J, Willliams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J, Conesa A (2008) High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Research, 36:3420-3435.

35. Götz S, Arnold R, Sebastian-Leon P, Martin-Rodriguez S, Tischler P, Jehl M-A, Dopazo J, Rattei T, Conesa A (2011) B2G-FAR, a species centered GO annotation repository. Bioinformatics, 27:919-924.

36. Grobler J, Bauer F, Subden RE, Van Vuuren HJ (1995) The *mae*1 gene of *Schizosaccharomyces pombe* encodes a permease for malate and other C4 dicarboxylic acids. Yeast, 11:1485-1491.

37. Hamilton CE, Gundel PE, Helander M, Saikkonen K (2012) Endophytic mediation of reactive oxygen species and antioxidant activity in plants: a review. Fungal Diversity, 4:1-10.

38. Heller J, Tudzynski P (2011) Reactive oxygen species in phytopathogenic fungi: signaling, development, and disease. Annual Review of Phytopathology, 49:369-390.

39. Hibbett DS, Binder M, Bischoff JF, Blackwell M, Cannon PF, Eriksson OE, Huhndorf S, James T, Kirk PM, Lücking R, Thorsten LH, Lutzoni F, Matheny PB, McLaughlin DJ, Powell MJ, Redhead S, Schoch CL, Spatafora JW, Stalpers JA, Vilgalys R, Aime MC, Aptroot A, Bauer R, Begerow D, Benny GL, Castlebury LA, Crous PW, Dai Y-C, Gams W, Geiser DM, Griffith GW, Gueidan C, Hawksworth DL, Hestmark G, Hosaka K, Humber RA, Hyde KD, Ironside JE, Kõljalg U, Kurtzman CP, Larsson K-H, Lichtwardt R, Longcore J, Miądlikowska J, Miller A, Moncalvo J-M, Mozley-Standridge S, Oberwinkler F, Parmasto E, Reeb V, Rogers JD, Roux C, Ryvarden L, Sampaio JP, Schüssler A, Sugiyama J, Thorn RG, Tibell L, Untereiner WA, Walker C, Wang Z, Weir A, Weiss M, White MM, Winka K, Yao Y-J, Zhang N (2007) A higher-level phylogenetic classification of the Fungi. Mycological Research, 111:509-547.

40. Hinton DM, Bacon CW (1985) The distribution and ultrastructure of the endophyte of toxic tall fescue. Canadian Journal of Botany, 63:36-42.

41. Hong LZ, Li J, Schmidt-Kuntzel A, Warren WC, Barsh GS (2011) Digital gene expression for non-model organisms. Genome Research, 21:1905-1915.

42. Jantasuriyat C, Gowda M, Haller K, Hatfield J, Lu G, Stahlberg E, Zhou B, Li H, Kim HR, Yu Y, Dean RA, Wing RA, Soderlund C, Wang G-L (2005) Large-scale identification of expressed sequence tags involved in rice and rice blast fungus interaction. Plant Physiology, 138:105-115.

43. Johnson-Cicalese J, Secks ME, Lam CK, Meyer WA, Murphy JA, Belanger FC (2000) Cross species inoculation of Chewings and strong creeping red fescue with fungal endophytes. Crop Science, 40:1485-1489.

44. Johnson L, Koulman A, Lane G, Fraser K, Voisey C, Pratt J, Bryan G, Johnson RD (2010) Multiple cyclic oligopeptides are synthesized by *Epichloë* endophytes via a single ribosomally encoded gene, *gigA*. Inoculum, 61:58.

45. Johnson LJ, Johnson RD, Schardl CL, Panaccione DG (2003) Identification of differentially expressed genes in the mutualistic association of tall fescue with *Neotyphodium coenophialum*. Physiological and Molecular Plant Pathology, 63:305-317.

46. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nature Review Genetics, 9:605-618.

47. Kuldau G, Bacon C (2008) Clavicipitaceous endophytes: their ability to enhance resistance of grasses to multiple stresses. Biological Control, 46:57-71.

48. Kutil BL, Greenwald C, Liu G, Spiering MJ, Schardl CL, Wilkinson HH (2007) Comparison of loline alkaloid gene clusters across fungal endophytes: predicting the co-regulatory sequence motifs and the evolutionary history. Fungal Genetics and Biology, 44:1002-1010.

49. Lam CK, Belanger FC, White JF Jr, Daie J (1994) Mechanism and rate of sugar uptake by *Acremonium typhinum*, an endophytic fungus infecting *Festuca rubra*: evidence for presence of a cell wall invertase in endophytic fungi. Mycologia, 86:408-415.

50. Leandro MJ, Goncalves P, Spencer-Martins I (2006) Two glucose/xylose tranporter genes from the yeast *Candida intermedia*: first molecular characterization of a yeast xylose-H+ symporter. Biochemical Journal, 395:543-549.

51. Li C, Nan Z, Li F (2008) Biological and physiological characteristics of *Neotyphodium gansuense* symbiotic with *Achnatherum inebrians*. Microbiological Research, 163:431-440.

52. Life Technologies, Thermo Fisher Scientific (2009) SAGE kit for SOLiD Next-generation sequencing. Retrieved from: http://www.lifetechnologies.com/us/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing/solid-next-generation-sequencing-sample-preparation-library-generation/sample-preparation-purification-solid-next-generation-sequencing/library-construction-solid-next-generation-sequencing/rna-library-construction-reagents-solid-next-generation-sequencing/sage-kit-solid-next-generation-sequencing.html

53. Lingner U, Munch S, Deising HB, Sauer N (2011) Hexose transporters of a hemibiotrophic plant pathogen. Functional variations and regulatory differences at different stages of

infection. Journal of Biological Chemistry, 286:20913-20922.

54. Lister R, Gregory BD, Ecker JR (2009) Next is now: new technologies for sequencing of genomes, transcriptomes, and beyond. Current Opinion in Plant Biology, 12:107-118.

55. Lutzoni F, Kauff F, Cox CJ, McLaughlin D, Celio G, Dentinger B, Padamsee M, Hibbett D, James TY, Baloch E, Grube M, Reeb V, Hofstetter V, Schoch C, Arnold AE, Miadlikowska J, Spatafora J, Johnson D, Hambleton S, Crockett M, Shoemaker R, Sung G-H, Lücking R, Lumbsch T, O'Donnell K, Binder M, Diederich P, Ertz D, Gueidan C, Hansen K, Harris RC, Hosaka K, Lim Y-W, Matheny B, Nishida H, Pfister D, Rogers J, Rossman A, Schmitt I, Sipman H, Stone J, Sugiyama J, Yahr R, Vilgalys R (2004) Assembling the fungal tree of life: progress, classification, and evolution of subcellular traits. American Journal of Botany, 91:1446-1480.

56. Marshall E, Costa LM, Gutierrez-Marcos J (2011) Cysteine-rich peptides (CRPs) mediate diverse aspects of cell-cell communication in plant reproduction and development. Journal of Experimental Botany, 62:1677-1686.

57. Marshall R, Kombrink A, Motteram J, Loza-Reyes E, Lucas J, Hammond-Kosack KE, Thomma BPHJ, Rud JJ (2011) Analysis of two in planta expressed LysM effector homologs from the fungus *Mycosphaerella graminicola* reveals novel functional properties and varying contributions to virulence on wheat. Plant Physiology, 156:756-769.

58. Marx F (2004) Small, basic antifungal proteins secreted from filamentous ascomycetes: a comparative study regarding expression, structure, function and potential application. Applied Microbiology and Biotechnology, 65:133-142.

59. Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, Kahl G, Reuter M, Krüger DH, Terauchi R (2003) Gene expression analysis of plant host–pathogen interactions by SuperSAGE. Proceedings of the National Academy of Sciences, USA 100:15718–15723.

60. Matsumura H, Ito A, Saitoh H, Winter P, Kahl G, Reuter M, Krüger DH, Terauchi R (2005) SuperSAGE. Cellular Microbiology, 7:11–18.

61. Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, Albertyn Z, Barrero RA, Krüger DH, Kahl G, Schroth GP, Terauchi R (2010) High-throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. PLoS ONE, 5:e12010.

62. Mentlak TA, Kombrink A, Shinya T, Ryder LS, Otomo I, Saitoh H, Terauchi R, Nishizawa Y, Shibuya N, Thomma BPHJ, Talbot NJ (2012) Effector-mediated suppression of chitin-triggered immunity by *Magnaporthe oryzae* is necessary for rice blast disease. Plant Cell, 24:322-335.

63. Meyer V (2008) A small protein that fights fungi: AFP as a new promising antifungal agent of biotechnological value. Applied Microbiology and Biotechnology, 78:17-28.

64. Meyer S, De Angeli A, Fernie AR, Martinoia E (2010) Intra- and extra-cellular excretion of carboxylates. Trends in Plant Science, 15:40-47.

65. Moon CD, Guillaumin J-J, Ravel C, Li C, Craven KD, Schardl CL (2007) New *Neotyphodium* endophyte species from the grass tribes Stipeae and Meliceae. Mycologia, 99:895-905.

66. Moy M, Li H., Sullivan R, White JF Jr, Belanger FC (2002) Endophytic fungal ß-1,6-glucanase expression in the infected host grass. Plant Physiology, 130:1298-1308.

67. Pang T-L, Chen F-C, Weng Y-L, Liao H-C, Yi Y-S, Ho C-L, Lin C-H, Chen M-Y (2010) Costars, a *Dictyostelium* protein similar to the C-terminal domain of STARS, regulates the actin cytoskeleton and motility. Journal of Cell Science, 123:3745-3755.

68. Plett JM, Martin F (2012) Mutualistic effectors: architects of symbiosis. Pages 295-326 in: Effectors in Plant-Microbe Interactions. Martin F, Kamoun S, eds. John Wiley & Sons, Inc., Oxford, UK.

69. Poole RL, Barker GLA, Werner K, Biggi GF, Coghill J, Gibbings JG, Berry S, Dunwell JM, Edwards KJ (2008) Analysis of wheat SAGE tags reveals evidence for widespread antisense transcription. BMC Genomics, 9:475.

70. Pratt LH, Liang C, Shah M, Sun F, Wang H, Reid SP, Gingle AR, Paterson AH, Wing R, Dean R, Klein R, Nguyen HT, Ma H-M, Zhao X, Morishige DT, Mullet JE, Cordonnier-Pratt M-M (2005) Sorghum expressed sequence tags identify signature genes for drought, pathogenesis, and skotomorphogenesis from a milestone set of 16,801 unique transcripts. Plant Physiology, 139:869-884.

71. Quéré R, Manchon L, Lejeune M, Clément O, Pierrat F, Bonafoux B, Commes T, Piquemal D, Marti J (2004) Mining SAGE data allows large-scale, sensitive screening of antisense transcript expression. Nucleic Acids Research, 32:e163.

72. Rasmussen S, Parsons AJ, Fraser K, Xue H, Newman JA (2008) Metabolic profiles of *Lolium perenne* are differentially affected by nitrogen supply, carbohydrate content, and fungal endophyte infection. Plant Physiology, 146:1440-1453.

73. Rasmussen S, Liu Q, Parsons AJ, Xue H, Sinclair B, Newman JA (2012) Grass-endophyte interactions: a note on the role of monosaccharide transport in the *Neotyphodium lolii-Lolium perenne* symbiosis. New Phytologist, 196:7-12.

74. Reddy PV, Lam CK, Belanger FC (1996) Mutualistic fungal endophytes express a protease which is homologous to proteases considered important in fungal pathogenicity. Plant Physiology, 111:1209-1218.

75. Regenberg B, Düring-Olsen L, Kiellnad-Brandt MC, Holmburg S (1999) Substrate specificity and gene expression of the amino-acid permeases in *Saccharomyces cerevisiae*. Current Genetics, 36:317-328.

76. Ren Q, Chen K, Paulsen IT (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. Nucleic Acids Research, e35:274-279.

77. Rep M (2005) Small proteins of plant-pathogenic fungi secreted during host colonization. FEMS Microbiology Letters, 253:19-27.

78. Robinson SJ, Cram DJ, Lewis CT, Parkin IAP (2004) Maximizing the efficacy of SAGE analysis identifies novel transcripts in *Arabidopsis*. Plant Physiology, 136:3223-3233.

79. Rodriguez RJ, White JF Jr, Arnold AE, Redman RS (2009) Fungal endophytes: diversity and functional roles. New Phytologist, 182: 314–330.

80. Rouxel T, de Wit JGM (2012) Dothideomycete effectors facilitiating biotrophic and necrotrophic lifestyles. Pages 195-218 in: Effectors in plant-microbe interactions. Martin F, Kamoun S, eds. John Wiley & Sons, Inc., Oxford, UK.

81. Rusk N, Kiermer V (2008) Primer: sequencing - the next generation. Nature Methods, 5:15.

82. Schardl CL (2001) *Epichloë festucae* and related mutualistic symbionts of grasses. Fungal Genetics and Biology, 33:69-82.

83. Schardl CL, Scott B, Florea S, Zhang D (2009) *Epichloë* endophytes: clavicipitaceous symbionts of grasses. Pages 275-306 in: Plant Relationships, 2nd Edition, The Mycota V. Deising H, ed. Springer-Verlag, Berlin Heidelberg.

84. Schardl CL, Young CA, Faulkner JR, Florea S, Pan J (2012) Chemotypic diversity of epichloae, fungal symbionts of grasses. Fungal Ecology, 5:331-344.

85. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk JW, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J, Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach J, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z. (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the *Clavicipitaceae* reveals dynamics of alkaloid loci. PLoS Genetics, 9(2):e1003323.

86. Scheller HV, Ulvskov P (2010) Hemicelluloses. Annual Review of Plant Biology, 61:263-289.

87. Scott B, Becker Y, Becker M, Cartwright G (2012) Morphogenesis, growth, and development of the grass symbiont *Epichloë festucae*. Pages 243-264 in: Morphogenesis and Pathogenicity in Fungi. Perez-Martin J, Di Pietro A, eds. Springer-Verlag, Berlin.

88. Scott B, Eaton CJ (2008) Role of reactive oxygen species in fungal cellular differentiations. Current Opinion in Microbiology, 11:488-493.

89. Siebert S, Robinson MD, Tintori S, Goeetz F, Helm RR, Smith SA, Shaner N, Haddock SHD, Dunn CW (2011) Differential gene expression in the siphonophore *Nanomia bijuga* (Cnidaria) assessed with multiple next-generation sequencing workflows. PLoS ONE 6:e22953.

90. Smarda P, Bures P, Horova L, Foggi B, Rossi G (2007) Genome size and GC content evolution of *Festuca*: ancestral expansion and subsequent reduction. Annals of Botany, 101:421-433.

91. Soanes DM, Alam I, Cornell M, Wong HM, Hedeler C, Paton NW, Rattray M, Hubbard SJ, Oliver SG, Talbot NJ (2008) Comparative genome analysis of filamentous fungi reveals gene family expansions associated with fungal pathogenesis. PLoS ONE 3:e2300.

92. Spiering MJ, Greer DH, Schmid J (2006) Effects of the fungal endophyte, *Neotyphodium lolii*, on net photosynthesis and growth rates of perennial ryegrass (*Lolium perenne*) are independent of *in planta* endophyte concentration. Annals of Botany, 98:379-387.

93. Sterck L, Rombauts S, Vandepoele K, Rouze P, Van de Peer Y (2007) How many genes are there in plants (… and why are they there)? Current Opinion in Plant Biology, 10:199-203.

94. Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland.

95. Tadych M, Ambrose KV, Bergen MS, Belanger FC, White JF Jr (2012) Taxonomic placement of *Epichloë poae* sp. nov. and horizontal dissemination to seedlings via conidia. Fungal Diversity, 54:117-131.

96. Tanaka A, Christensen MJ, Takemoto D, Park P, Scott B (2006) Reactive oxygen species play a role in regulating a fungus-perennial ryegrass mutualistic association. Plant Cell, 18:1052-1066.

97. Tanaka A, Takemoto D, Hyon G-S, Park P, Scott B (2008) NoxA activation by the small GTPase RacA is required to maintain a mutualistic symbiotic association between *Epichloë festucae* and perennial ryegrass. Molecular Microbiology, 68:1165-1178.

98. Tanaka A, Takemoto D, Chujo T, Scott B (2012) Fungal endophytes of grasses. Current Opinion in Plant Biology, 15:462-468.

99. Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B (2005) A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. Molecular Microbiology, 57:1036-1050.

100. Teichert S, Rutherford JC, Wottawa M, Heitman J, Tudzynski B (2008) Impact of ammonium permeases MepA, MepB, and MepC on nitrogen-regulated secondary metabolism in *Fusarium fujikuroi*. Eukaryotic Cell, 7:187-201.

101. Templeton MD, Rikkerink EHA, Beever RE (1994) Small, cysteine-rich proteins and recognition in fungal-plant interactions. Molecular Plant-Microbe Interaction, 7:320-325.

102. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25:4876-4882.

103. Torres MA (2010) ROS in biotic interactions. Physiologia Plantarum, 138:414-429.

104. Velculescu V, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science, 270:484–487.

105. Werner A, Swan D (2010) What are natural antisense transcripts good for? Biochemical Society Transactions, 38:1144-1149.

106. White JF Jr, Torres MS (2010) Is plant endophyte-mediated defensive mutualism the result of oxidative stress protection? Physiologia Plantarum, 138:440-446.

107. Young CA, Bryant MK, Christensen MJ, Tapper BA, Bryan GT, 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.

108. Young CA, Felitti S, Shields K, Spangenberg G, Johnson RD, Bryan GT, Saikia S, Scott B (2006) A complex gene cluster for indole-diterpene biosynthesis in the grass endophyte *Neotyphodium lolii*. Fungal Genetics and Biology, 43:679-693.

109. Yue Q, Johnson-Cicalese J, Gianfagna TJ, Meyer WA (2000) Alkaloid production and chinch bug resistance in endophyte-inoculated Chewings and strong creeping red fescues. Journal of Chemical Ecology, 26:279-292.

110. Zhang N, Zhang S, Borchert S, Richardson K, Schmid J (2011) High levels of a fungal superoxide dismutase and increased concentration of a PR-10 plant protein in associations between the endophytic fungus *Neotyphodium lolii* and ryegrass. Molecular Plant-Microbe Interaction, 24:984-992.

Chapter 3

Horizontal gene transfer of a bacterial insect toxin gene into the *Epichloë* fungal symbionts of grasses

Chapter 3

Horizontal gene transfer of a bacterial insect toxin gene into the *Epichloë* fungal symbionts of grasses

Introduction

Horizontal gene transfer (HGT), also known as lateral gene transfer involves movement of genetic material between unrelated organisms and differs from the usual parent to offspring vertical transmission of genes (Keeling and Palmer, 2008). Such DNA transmissions, although common among prokaryotes, are rare when the donor and recipient belong to different domains or kingdoms (Blaxter, 2007). Reports of HGT events across various taxonomic ranks have been more forthcoming in recent years due to the massive advances in DNA sequencing technologies (Fitzpatrick et al., 2008; Yoshida et al., 2010; Moran et al., 2010; Kristiansson et al., 2011; Richards et al., 2011; McDonald et al., 2012; Moran et al., 2012; Xi et al., 2012; Li et al., 2013; Zhang et al., 2014).

In some cases, HGT events facilitate the acquisition of adaptive evolutionary traits that are deemed highly beneficial to the recipient. One such transmission, exemplified by the lateral movement of a bacterial mannanase gene into the insect *Hypothenemus hampei*, the coffee berry borer beetle, is hypothesized to have led to the adaptation of the insect to a new host, the coffee berry, which is rich in mannans (Acuña et al., 2012). Another study reported that recurrent HGT events involving the plant nuclear genes encoding C_4 enzymes phosphoenolpyruvate carboxylase (PEPCase) and phosphoenolpyruvate carboxykinase (PEPCK) from various grass lineages led to the C_4 photosynthetic adaptation of some members of the grass genus *Alloteropsis* (Christin et al., 2012). C_4 photosynthesis is advantageous to plants living in warm habitats, where the rate of photorespiration is elevated, because the efficiency of water and nutrient utilization is increased. The photosynthetic efficiency is driven by the concentration of carbon dioxide around the

ribulose-1,5-bisphosphate carboxylase oxygenase enzyme (RuBisCo) with the aid of PEPCase and PEPCK (Sage, 2004). Hence, the HGT acquisition of these carbon fixing and decarboxylating enzymes by several *Alloteropsis* spp. is considered to be beneficial in adapting to arid conditions.

Epichloë-infected turf grasses are known to have anti-insect properties that are due to the production of toxic fungal alkaloids by the endophytes *in planta* (Schardl, 2001; Schardl et al., 2013). Previous studies have demonstrated that the presence of *E. festucae* endophyte in *Festuca rubra* provided resistance to hairy cinch bug insect (*Blissus leucopterus hirtus*), a damaging turf grass pest (Saha et al., 1987; Yue et al., 2000). The phenomenon of insect resistance in several *Festuca rubra* turf subspecies was first observed in field trials designed to explore the utility of endophytes in improving turf grass performance in the early 1980s (Saha et al., 1987). Other studies since then have linked the protection against insect herbivory to the loline and peramine alkaloids produced by the endophytes living symbiotically within the host turf grass (Tanaka et al., 2005; Spiering et al., 2007; Schardl et al., 2007).

In this chapter, the discovery and functional characterization of a horizontally transferred gene of bacterial origin into the *Epichloë* endophytes is described. This gene is similar to the well-studied *Photorhabdus luminescens makes caterpillars floppy* (*mcf1* and *mcf2*) (Daborn et al., 2002; Waterfield et al., 2003) and *Pseudomonas fluorescens fitD* (Pechy-Tarr et al., 2008) insect toxin encoding genes. The genes contain the highly conserved TcdA/TcdB pore-forming domain (Pfam12920) (Punta et al., 2012) similar to that of *Clostridium difficile* toxin B (Genisyuerek et al., 2011). *mcf1* has a 1,623 bp N-terminal extension relative to *mcf2* and all three bacterial genes encode large proteins (2,388 to 2,996 amino acids).

Ph. luminescens is a symbiotic bacterium of entomopathogenic nematodes of the family Heterorhabditidae. When the nematodes infect their insect host, the bacteria are regurgitated into

the insect hemolymph where they release multiple toxins that kill the insect host (ffrench-Constant et al., 2007; Hinchliffe et al., 2010). The *mcf1* gene was discovered in a screen of *Ph. luminescens* cosmid clones for insect toxicity, and the name comes from the effect seen after normally non-toxic *Escherichia coli* expressing the gene were injected into caterpillars (Daborn et al., 2002). *mcf2*, also from *Ph. luminescens*, and *fitD*, from *Pseudomonas protegens* Pf-5 and CHA0 isolates (previously called *Ps. fluorescens*, Ramette et al., 2011) were discovered by their sequence similarity to *mcf1* and have been shown to have a similar effect on caterpillars (Waterfield et al., 2003; Pechy-Tarr et al., 2008). Mcf1 was found to kill insects by promoting apoptosis via the mitochondrial pathway (Dowling et al., 2004; ffrench-Constant et al., 2007). The *Ps. protegens* FitD toxin contributes to the oral insecticidal activity of the bacteria (Ruffner et al., 2013). The mode of action of Mcf2 is unknown (Waterfield et al., 2003).

Interestingly, of the 568 fungal genome sequences currently available at NCBI (as of March 2014), only the *Epichloë* have a *mcf-like* gene, suggesting a lineage-specific HGT event from bacteria into a predecessor of the modern *Epichloë*. The *mcf-like* gene in *E. typhina* subsp. *poae* infecting *Poa secunda* subsp. *juncifolia* (big blue grass), *E. typhina* subsp. *poae* Ps1 (Tadych et al., 2012; Leuchtmann et al., 2014) is expressed and functional, suggesting that it may be a component, in addition to the fungal alkaloids, in the observed insect resistance of some endophyte-infected grasses.

The work discussed here is published in Ambrose KV, Koppenhöfer AM, Belanger FC (2014) Horizontal gene transfer of a bacterial insect toxin gene into the *Epichloë* fungal symbionts of grasses. Scientific Reports, 4:5562.

Materials and methods

Plant and fungal samples

Epichloë spp.-infected plants were grown by clonal propagation and maintained in 15 cm pots in the greenhouse. Each *Epichloë* sp. was isolated from its host plant by plating surface-sterilized leaf sheath tissue on potato dextrose agar (Difco Laboratories, Detroit, MI).

DNA and RNA isolation

Genomic DNA of *Epichloë* spp. was extracted from culture grown in potato dextrose broth for 14 days on a shaker (175 rpm) at room temperature (23-25°C). The DNA was isolated as previously described (Moy et al., 2002). RNA was obtained from the innermost leaf sheath tissues of *Epichloë*-infected plants. RNA isolation was as previously described (Ambrose and Belanger, 2012). Nucleic acid concentration was measured using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

Amplification and sequencing of *Epichloë mcf-like* genes and cDNAs

Initial screening to determine the presence of the *mcf-like* gene in *E. festucae* RC and *E. typhina* subsp. *poae* Ps1 fungal isolates was done using fungal genomic DNA. Upon detection of the gene, first-strand cDNA from 5 µg *Epichloë*-infected plant total RNA was synthesized by using SuperScript™ III Reverse Transcriptase (Life Technologies, Carlsbad, CA) and 500 ng of oligo$(dT)_{18}$ primer according to the manufacturer's instructions. PCR was performed in 50 µL with either 0.5 µg of fungal genomic DNA or 1 µL *Epichloë*-infected plant cDNA generated from oligo$(dT)_{18}$, 1.25 units of PrimeSTAR HS DNA Polymerase, 1X PrimeSTAR Buffer with Mg^{2+} (TaKaRa Bio Inc., Shiga, Japan), 200 µM each dNTP, and 0.3 µM of each forward and reverse primer (Integrated DNA Technologies, Inc., Coralville, IA). Two-step PCR was performed by template denaturation at 98°C for 10 seconds followed by 6 minutes extension at 68°C for 30 cycles in a GeneAmp 9700 thermal cycler (Applied Biosystems, Inc., Foster City, CA). The

amplification products were visualized on a 1% TBE agarose gel. The amplified genomic and cDNA PCR fragments were purified by using 0.5X Agencourt AMPure XP (Beckman Coulter, Inc., Brea, CA) to remove any fragments under 1 kb and then sequenced directly (Genewiz, Inc., South Plainfield, NJ). For each sequencing reaction, about 800 ng of purified PCR product in 10 µL was treated with 2 µL ExoSAP-IT (USB Corp., Cleveland, OH) to remove unincorporated primers and excess dNTPs. The ExoSAP-IT reaction was performed at 37°C for 15 min followed by heating at 80°C for 15 minutes to inactivate the enzymes.

Expression of the *E. typhina* subsp. *poae* Ps1-*mcf* gene in *E. coli*

To test the anti-insect activity of the *E. typhina* subsp. *poae* Ps1-Mcf protein, the cDNA was cloned into the expression vector pCold II (TaKaRa Bio). The *E. typhina* subsp. *poae* Ps1 cDNA was amplified by PCR as described above with oligonucleotides that introduced a *Kpn*I site at the 5' end and an *Eco*RI site at the 3' end. The sequences of the oligonucleotides are: forward 5'-GATATAACCATGGCTCACAACACT-3' and reverse 5'-TCAACTGAATTCCTACTGATTTCCAGC-3'. Two hundred ng of the purified PCR product was digested with the restriction enzyme *Kpn*I (TaKaRa Bio), purified with 0.35X Agencourt AMPure XP to remove any fragments under 1.5 kb, digested with *Eco*RI, and again purified with 0.35X AMPure XP. The expression vector pCold II was similarly digested with *Kpn*I and *Eco*RI. The digested pCold II plasmid was treated with shrimp alkaline phosphatase (Affymetrix, Santa Clara, CA) to prevent vector religation. Overnight ligation of the digested *E. typhina s*ubsp. *poae* Ps1-*mcf* PCR fragment and pCold II plasmid was done at a 3:1 insert: vector molar ratio using T4 DNA ligase (New England Biolabs, Inc., Ipswich, MA).

Two µL of the ligation product was used to transform 20 µL *E. coli* DH5α electroporation-competent cells. The transformed cells were incubated in SOC medium for 1 hour at 37°C, followed by overnight growth of cells on LB medium supplemented with 100 µg/mL ampicillin.

Transformed bacterial colonies were screened for plasmids containing the *E. typhina* subsp. *poae* Ps1-*mcf* gene insert by using PCR as described above. Selected clones were grown in 3 mL of LB + ampicillin broth overnight followed by plasmid purification using QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). DNA sequencing of the plasmids were done as described above. A plasmid containing the correct sequence was transformed into *E. coli* BL21-CodonPlus (DE)-RIPL competent cells (Agilent Technologies, Santa Clara, CA). For use as a control, a transformant containing the pCold II vector only was also generated.

For expression of the *E. typhina* subsp. *poae* Ps1-Mcf protein, the BL21-CodonPlus (DE3)-RIPL transformant was induced by cold shock and isopropyl-β-D-1-thiogalactopyranoside (IPTG). Loss of *E. typhina* subsp. *poae* Ps1-*mcf*::pCold II recombinant plasmid in *E. coli* cells were detected during the course of long induction periods. This was not observed in *E. coli* cells containing pCold II vector only. Previous studies have shown that carbenicillin and ampicillin, even in high concentrations, quickly lose their ability to maintain selective pressure due to saturation of the media with the antibiotic degrading enzyme β-lactamase (Collins et al., 2013; Korpimaki et al., 2003). Extra metabolic burden placed on the bacterial cells to maintain the large recombinant plasmid combined with the steady decrease of β-lactam in culture may have contributed to the plasmid loss that was observed. In order to circumvent this problem, the culture supernatant was replaced with fresh medium at regular intervals.

To induce expression a 10 mL culture of the cells was grown for 15 hours at 30°C in LB and 200 µg/mL carbenicillin. A 400 µL aliquot of the culture was pelleted, resuspended in 400 µL, added to 10 mL of fresh LB plus carbenicillin and incubated at 30°C to an OD of 0.4-0.5. The culture was then cold-shocked for 15 minutes on ice and subsequently pelleted in a Model J2-21 centrifuge (Beckman Coulter, Inc., Brea, CA) set to 5°C to obtain the bacterial cells. Fresh LB broth supplemented with carbenicillin was used to resuspend the bacterial pellets. To induce protein expression in the culture, IPTG was added to 0.1 mM and the culture was then incubated

at 15°C with rotational shaking for 21-24 hours. The bacterial cells were pelleted and resuspended in fresh LB and carbenicillin every 3 hours for the initial 6 hours and again after 12 hours to prevent β-lactamase saturation in the culture supernatant. The transformant containing pCold II vector only was treated the same way.

For sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis, 1 mL each of *E. coli* cells expressing the *E. typhina* subsp. *poae* Ps1-Mcf protein and the vector only control cultures were pelleted and lysed using 1X FastBreak Cell Lysis Reagent (Promega, Madison, WI) following the manufacturer's protocol. Following lysis, samples were centrifuged at 10,000 RPM for 15 minutes and the supernatants removed. The proteins remaining in the pellet were solubilized by adding 100 µL of 2X SDS sample buffer (Laemmli, 1970), boiled for 5 minutes, and 30 µL was then subjected to electrophoresis on a 10% gel.

Insect assays to determine toxicity of *E. typhina* subsp. *poae* Ps1-Mcf protein

BL21-CodonPlus (DE3)-RIPL cells expressing the *E. typhina* subsp. *poae* Ps1-Mcf::pCold II recombinant insert and vector only control were prepared as described above. Cells were pelleted and resuspended in sterile water. The insect assays were carried out by Dr. Albrecht Koppenhöfer. Black cutworm, *Agrotis ipsilon* (Hufnagel), eggs were obtained from Benzon Research Inc. (Carlisle, PA) and reared on insect diet (Southland Products Inc. Lake Village, AR) at 27 °C and 14:10 (light:dark) hours to fourth instars. Immediately before injection the larvae were surface sterilized in 0.8% sodium hypochloride for 30 seconds and then rinsed in sterile distilled water. Six to ten fourth instars were injected with 5 x 10^7 *E. coli* cells containing either the *E. typhina* subsp. *poae* Ps1-Mcf::pCold II induced protein or the pCold II vector only control. Additional controls were water injections and untreated samples. The assay was conducted twice. Injections were in a total volume of 5 µL using a Hamilton microsyringe with a 30-gauge needle. The larvae were then placed individually in 30 mL plastic cups with 10 g of moist (10% w/v)

pasteurized (3 hours at 72 °C) sand with about 1 cm^2 diet and incubated at temperature (22-24 °C). Mortality was assessed at 24, 48, 72, and 120 hours after injection. Larvae were considered dead when they no longer reacted to being poked repeatedly with a probe. The data at the final time points were analyzed by Fisher's exact test to compare the toxin-expressing samples to the vector only controls. Significance was evaluated at $P < 0.05$.

Schematic representation of *mcf* gene structure

Schematic representation of the *mcf* gene structures was done using Gene Structure Display Server (Guo et al., 2007; http://gsds.cbi.pku.edu.cn/index.php?input=site). Illustrative comparison of amino acid similarity for Mcf1, Mcf2 and *E. typhina* subsp. poae Ps1-Mcf protein was performed using Gene Evolution and Conservation Analysis (GECA) (Fawal et al., 2012; http://peroxibase.toulouse.inra.fr/tools/geca_input_demo.php).

Phylogenetic analysis

To identify genes in addition to MCM7 and TSR1 for use in the phylogenetic analyses, the FUNYBASE genome database (http://genome.jouy.inra.fr/funybase/) (Marthey et al., 2008) was searched for single copy orthologs identified as phylogenetically high-performing genes by Aguileta et al. (2008). Two genes with topological scores of 96 or above and that also had significant sequence polymorphism within *Epichloë* were identified, FG570 (NAD-dependent glutamate dehydrogenase) and MS444 (isoleucine tRNA synthase). DNA sequences for these genes were obtained from the Genome Project at the University of Kentucky website (Schardl et al., 2013) or from NCBI.

The Clustal-X program (Thompson et al., 1997) was used to align the DNA coding sequences. The alignments generated by Clustal-X were modified manually to minimize gaps. The phylogenetic analyses were performed with the PAUP* program, version 4.0b10 for Macintosh

(Swofford, 2002). The phylogenetic analyses were done by using the maximum parsimony full heuristic search option set to random sequence addition, tree-bisection-reconnection (TBR) branch swapping, and Multrees on, with 1,000 bootstrap replications. Gaps were treated as missing data. The DNA sequences were also analyzed by the maximum likelihood method in the PAUP* program, which generated trees of similar topology to those of the maximum parsimony analysis (not shown). For the maximum likelihood analyses, the trees were generated with a fast heuristic search using the HKY85 model of sequence evolution, and 100 bootstrap replications. The K_s and K_a values based on pairwise comparison between species were determined by using the MEGA5.2 program with the Nei-Gojobori, Jukes-Cantor model (Tamura et al., 2011).

Results

Detection of a *makes caterpillars floppy-like* gene in *Epichloë* genomes

The presence of an *Epichloë* gene that showed homology to the bacterial *mcf* insect toxin genes was discovered during a targeted screen for putative horizontally transferred genes in the *E. festucae* 2368 genome (Schardl et al., 2013). Genes that were flanked by mobile DNA elements such as transposons were selected for identification by BLASTx searches by queries against all available protein sequences to infer their identity (Altschul et al., 1990). The rationale behind the strategy was that since mobile genetic elements have been widely implicated in gene transfer within and between genomes (Syvanen, 1984; Frost et al., 2005; Cortez et al., 2009; Acuña et al., 2012), there could be a possibility of HGT facilitated by transposons into the *Epichloë* genome.

One candidate was observed to have 34-36% identity to the bacterial Mcf and FitD proteins. A tag similar to those bacterial genes was also recovered in our SOLiD-SAGE quantitative transcriptome expression data (Ambrose and Belanger, 2012). Analysis of 13 available *Epichloë* whole genome sequences (Schardl et al., 2013) revealed that the *mcf-like* gene is present in all of them and most species contained a gene with long open-reading frames interrupted by two introns

(Figure 3.1a).

The sizes of the *Epichloë mcf-like* gene are smaller than *mcf1* (8,790 bp), *mcf2* (7,167 bp) and *fitD* (8,991 bp) bacterial genes and ranged from 5,532 to 6,132 bp. The *Epichloë mcf-like* gene structures are similar throughout the genus in that an intron downstream of the conserved TcdA/TcdB pore-forming domain is consistently present. Additionally, there was another intron observed close to the N-terminal in all *Epichloë mcf-like* genes except in three *E. festucae* isolates. The *mcf-like* genes of *E. festucae* 2368 and *E. festucae* Fl1 are identical to each other while *E. festucae* Rose City's *mcf-like* gene was found have a mutation that resulted in a frame-shift. Hence the Rose City isolate is not expected to produce an intact full-length Mcf-like protein. The nucleotide insertion was determined experimentally since this isolate does not have a whole genome sequence available.

Examination of the *Epichloë mcf-like* genes also indicated early termination in three other species (*E. elymi, E. gansuensis* and *E. typhina* subsp. *poae* 5819) due to premature stop codons that prevented translation into full-length proteins containing the TcdA/TcdB pore-forming domain. The cDNA and genomic DNA of the *mcf-like* gene of *E. typhina* subsp. *poae* Ps1 (henceforth designated as *E. typhina* subsp. *poae* Ps1-*mcf*) were sequenced and its intron positions determined in order to compare the gene with its homologs in the annotated genomes of *E. typhina* subsp. *poae* 5819 and *E. typhina* E8. The gene structures in *E. gansuensis* var. *inebrians* and *E. typhina* E8 were unusual relative with the other *Epichloë* species in that these two isolates were annotated as having extended N-terminal sequences compared to most of the other species. In both cases the annotated extension can be attributed to a deletion of the presumed normal coding sequence start site and recognition of upstream open reading frames as exons. Sequence comparison of *E. typhina* subsp. *poae* Ps1-Mcf with the bacterial Mcf1 and Mcf2 showed remarkable similarity and conservation throughout the 1,997 amino acid long protein, and is especially striking in the

TcdA/TcdB pore forming domain (Figure 3.1b). Amino acid identities in the region of overlap were found to be very similar between the *Epichloë* Mcf-like sequences and all three bacterial proteins (33-36%; Table 3.1) making the identification of the donor bacterial species a challenge.

The lack of introns in the *mcf1*, *mcf2*, or *fitD* genes is characteristic of bacterial genes; however, the presence of presumed introns in all of the *Epichloë mcf-like* genes suggests that the introns seen in the fungal genes have been acquired after the HGT event, while simultaneously eliminating the possibility of bacterial contamination in the *Epichloë*-infected plant nucleic acid samples that was used in the study. Additionally, the identical placement of the introns in all *Epichloë* species hints at the HGT event occurring once into the most basal member of the lineage followed by vertical inheritance via speciation throughout the genus. As described earlier, all *Epichloë mcf-like* gene structures are similar in that they have an intron downstream of the TcdA/TcdB pore-forming domain and in most, another intron close to the N-terminal, although the intron sequences themselves varied from species to species. These identical intron positions give credence to the likelihood that there was a single HGT acquisition of the *mcf-like* gene into the original fungal recipient genome as well as a unique intron insertion event. Previous reports have detailed such intron acquisitions in other eukaryotic genes of bacterial origin (Marcet-Houben and Gabaldon, 2010; Da Lage et al., 2013). This phenomenon has been widely attributed to a process referred as amelioration (Lawrence and Ochman, 1997). In such events, laterally transferred genes are initially similar to the donor genome in terms of base composition, however selective pressure in the recipient genome results in the horizontally acquired genes to become similar to other genes in the new genome over time.

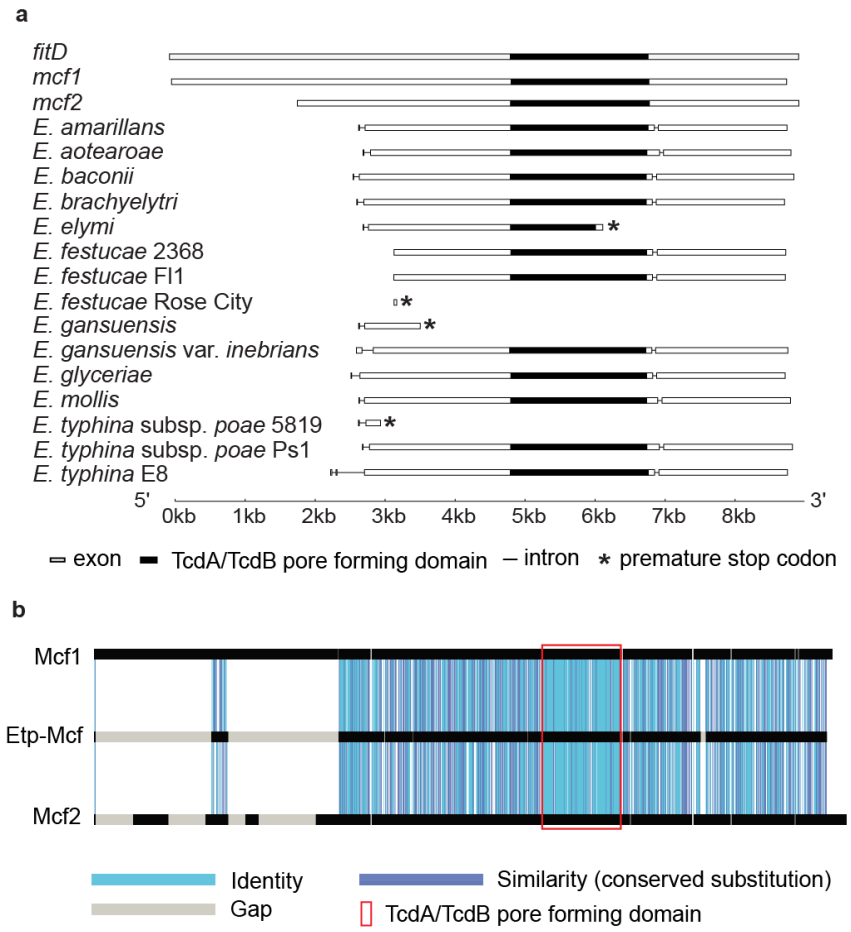


Figure 3.1. Gene structure of the *Epichloë mcf-like* genes and amino acid similarity of *E. typhina* subsp. *poae* Ps1-Mcf protein with the bacterial Mcf proteins. (**a**) Diagrams of gene structure of the bacterial *mcf1*, *mcf2*, *fitD*, and *Epichloë mcf-like* genes. The exons are indicated by boxes and the introns by lines. Filled in boxes indicate the conserved TcdA/TcdB pore-forming region. The positions of the premature stop codons in some of the *Epichloë* genes are indicated by *. The end of intron 1 of the *E. typhina* subsp. *poae* 5819 sequence was modified from the annotated version to that of the experimentally determined position in *E. typhina* subsp. *poae* Ps1. (**b**) GECA analysis illustrating the amino acid sequence similarity of the *E. typhina* subsp. *poae* Ps1-Mcf protein (1,997 amino acids; accession KJ502561) to Mcf1 (2,929 amino acids; accession AF503504.2) and Mcf2 (2,388 amino acids; accession AY445665) proteins from *Ph. luminescens*.

Table 3.1. Comparison of *Epichloë* Mcf-like amino acid sequences with those of the bacterial Mcf1, Mcf2, and FitD sequences.

Species	Identity		
	Mcf1 (AAM88787.1) 2,929 aa	Mcf2 (AAR21118.1) 2,388 aa	FitD (AFD97974.1) 2,996 aa
E. amarillans 1,994 aa	673/2,032 (33%)	649/1,856 (35%)	673/1,922 (35%)
E. aotearoae 1,986 aa	676/2,037 (33%)	629/1,846 (34%)	660/1,909 (35%)
E. baconii 2,044 aa	662/1,924 (34%)	652/1,861 (35%)	672/1,915 (35%)
E. brachyelytri 1,988 aa	636/1,839 (35%)	638/1,836 (35%)	658/1,912 (34%)
E. festucae 2368 1,844 aa	619/1,883 (33%)	631/1,881 (34%)	611/1,766 (35%)
E. festucae Fl1 1,844 aa	617/1,883 (33%)	631/1,881 (34%)	611/1,766 (35%)
E. gansuensis var. *inebrians* 1,983 aa	669/1,991 (34%)	677/1,997 (34%)	671/1,883 (36%)
E. glyceriae 2,009 aa	655/1,924 (34%)	639/1,842 (35%)	663/1,865 (36%)
E. mollis 2,012 aa	644/1,870 (34%)	657/1,850 (36%)	682/1,922 (35%)
E. typhina subsp. *poae* Ps1 1,997 aa	652/1,899 (34%)	673/1,985 (34%)	669/1,921 (35%)
E. typhina E8 2,004 aa	673/2,023 (33%)	660/1,955 (34%)	667/1,944 (34%)

Estimation of the age of the *mcf-like* gene HGT event into *Epichloë*

In order to gain an insight into the age of the *mcf-like* gene HGT event into *Epichloë,* phylogenetic trees were reconstructed using four genes that were identified as optimal in several recent studies involving a wide range of fungal taxa (Aguileta et al., 2008; Schmitt et al., 2009; Raja et al., 2011; Tretter et al., 2013). These four genes are MCM7 (a subunit of the mini chromosome maintenance complex; Alabert and Groth, 2011), TSR1 (a gene required for synthesis of 40S ribosomal subunits; Gelperin et al., 2001), NAD-dependent glutamate

dehydrogenase and isoleucine tRNA synthase.

Maximum parsimony phylogenetic analyses of MCM7, TSR1, NAD-dependent glutamate dehydrogenase, and isoleucine tRNA synthase sequences from 13 *Epichloë*, as well as some related Clavicipitaceae, are shown in Figures 3.2 and 3.3. Accession numbers of the sequences are given in Table 3.2. To facilitate the alignments for these analyses the intron sequences were removed and only the DNA coding sequences were used. The *Metarhizium anisopliae* sequences were designated as the outgroups for rooting the trees since *M. anisopliae* is in a sister clade, which diverged prior to that of *Epichloë* (Sung et al., 2008). In all of the trees, the two *E. gansuensis* sequences were in a sister clade to the rest of the *Epichloë* spp., which had *E. glyceriae* at its base. Both *E. gansuensis* and *E. glyceriae* carry the *mcf-like* gene in their genomes, thus the HGT event must have happened in the most common ancestor of these two basal *Epichloë* spp.

The age of the HGT event was estimated using a molecular clock approach (Li, 1997) similar to that used to estimate the age of polyploidy in cotton (Senchina et al., 2003), the divergence times of maize LTR-retrotransposons (Brunner et al., 2005), and the age of a segmental duplication in maize (Xu and Messing, 2008). The divergence rate for each gene was estimated by calibration to *Atkinsonella hypoxylon* using the formula $T = K_s/2r$ was first estimated. *A. hypoxylon* (synonym: *Balansia hypoxylon*) (Ginns, 2011) is closely related to *B. epichloë* (Kuldau et al., 1997). *B. epichloë* was placed in a sister clade to the clade containing *E. typhina* and the *Claviceps* spp. and the divergence of those two clades was estimated to have occurred 81 million years ago (Sung et al., 2008).

The K_s (substitutions at synonymous sites) and K_a (substitutions at nonsynonymous sites) for the coding sequence of each gene relative to the *A. hypoxylon* ortholog was calculated (Table 3.3). In

this analysis glutamate dehydrogenase from *E. gansuensis* var *inebrians* was not included, since the gene sequence was split between two contigs, which introduced a gap into the alignment. The low K_a/K_s ratios for these four genes indicate they are under purifying selection. From the K_s for each gene, the rate of divergence of synonymous sites (r) for each gene was calculated (Table 3.3). The calculated rates varied from 5.93×10^{-9} to 7.72×10^{-9} substitutions per synonymous site per year, which are similar to reported fungal gene divergence rates (Kasuga et al., 2002).

The calculated rates of divergence of each gene were then used to calculate the time of divergence (T) between the *Claviceps* clade and the *Epichloë* clade as well as between *E. glyceriae* and *E. gansuensis* sequences (Tables 3.4 and 3.5). The mean calculated divergence time between the *Claviceps* clade and the *Epichloë* clade was 58.8 million years ago (Mya) and the mean divergence time between *E. gansuensis* and *E. glyceriae* was 7.2 Mya. These estimates point to the single acquisition of the *mcf-like* gene into *Epichloë* sometime after the divergence of the *Claviceps* and *Epichloë* clades 58.8 Mya, and before the divergence of *E. gansuensis* and *E. glyceriae* 7.2 Mya.

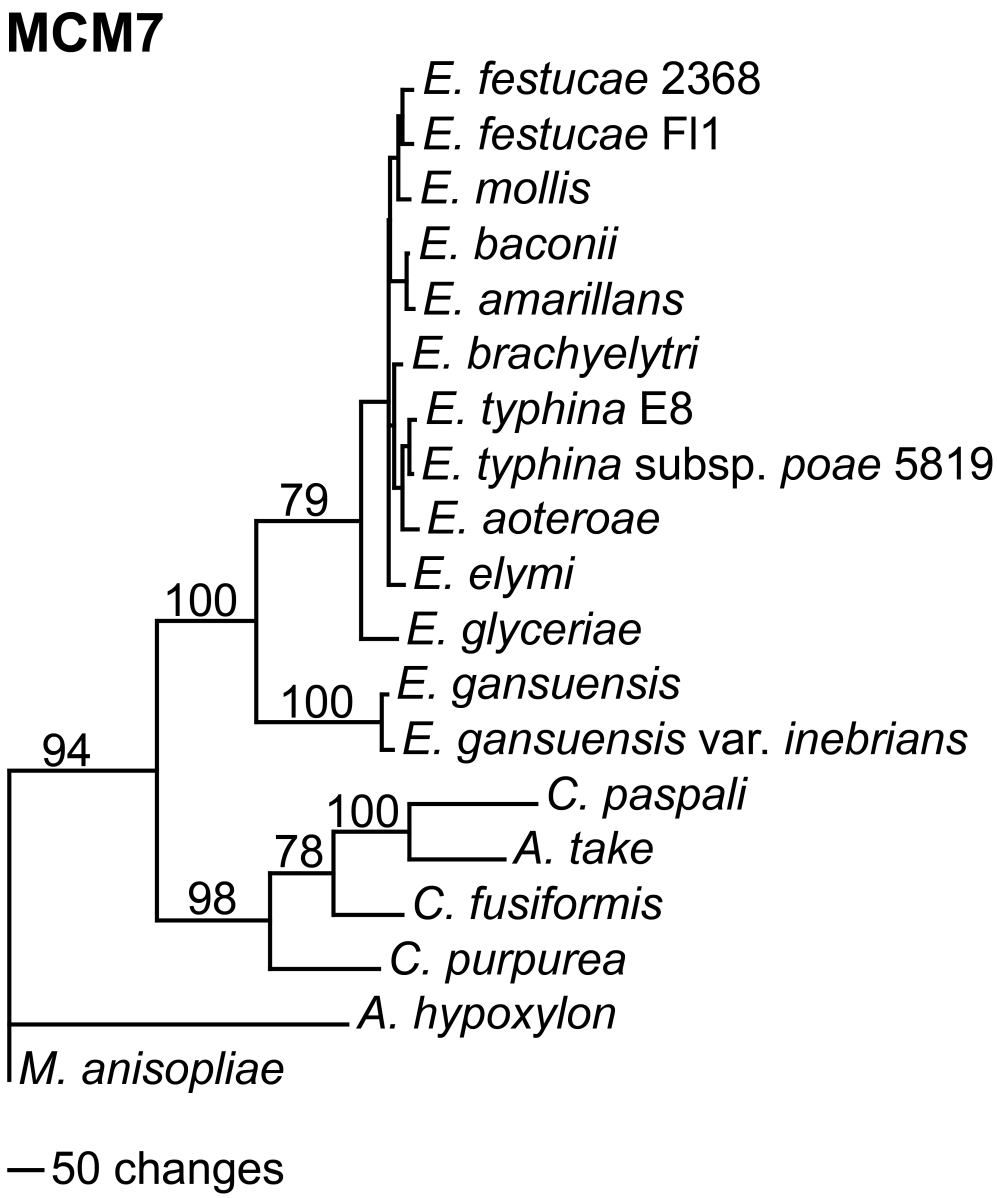


Figure 3.2. Rooted 50% majority rule maximum parsimony phylogenetic tree of the MCM7 DNA coding sequences. The *Metarhizium anisopliae* sequences were designated as the outgroup for rooting the tree. The numbers at the nodes are the bootstrap percentages based on 1,000 replications. The tree was based upon 2,472 total characters, of which 1,549 were constant, 343 variable characters were parsimony uninformative, and 580 variable characters were parsimony informative. The unmarked nodes all had bootstrap support of 69 or higher.

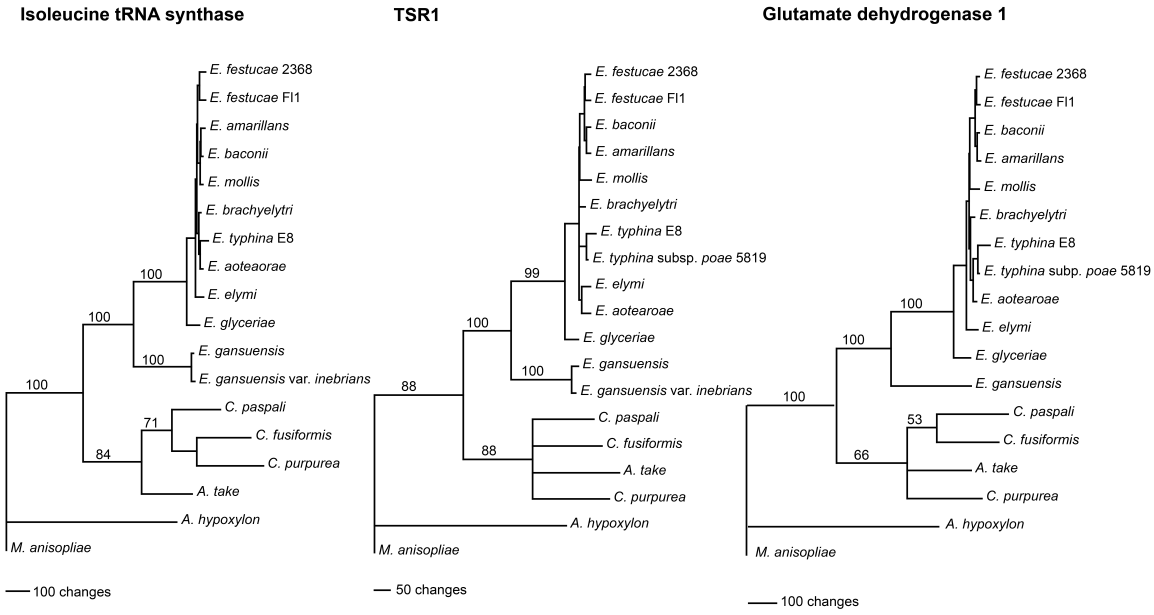


Figure 3.3. Rooted 50% majority rule maximum parsimony phylogenetic trees of the glutamate dehydrogenase, isoleucine tRNA synthase, and TSR1 DNA coding sequences. The *Metarhizium anisopliae* sequences were designated as the outgroup for rooting the trees. The numbers at the nodes are the bootstrap percentages based on 1,000 replications. The TSR1 tree was based upon 2538 total characters, of which 1433 were constant, 478 variable characters were parsimony uninformative, and 627 variable characters were parsimony informative. The unmarked nodes all had bootstrap support of 53 or higher. The glutamate dehydrogenase tree was based upon 3225 total characters, of which 1967 were constant, 493 variable characters were parsimony uninformative, and 765 variable characters were parsimony informative. The glutamate dehydrogenase sequence from *E. gansuensis* var *inebrians* was not included, since the gene sequence was split between two contigs, which introduced a gap into the alignment. The unmarked nodes all had bootstrap support of 66 or higher. The isoleucine tRNA synthase tree was based upon 3228 total characters, of which 1884 were constant, 517 variable characters were parsimony uninformative, and 827 variable characters were parsimony informative. The isoleucine tRNA synthase gene from *E. typhina* subsp. *poae* 5819 was not included since the gene

was split between two contigs, which introduced a gap into the alignment. The unmarked nodes all had bootstrap support of 70 or higher.

Table 3.2. Sources of DNA sequences used in this study.

Species	**Source**
Aciculosporium take	AFQZ00000000.1
Atkinsonella hypoxylon	http://www.endophyte.uky.edu/
Claviceps fusiformis	AFRA00000000.1
C. paspali	AFRC00000000.1
C. purpurea	CAGA00000000
Metarhizium anisopliae	ADNJ00000000.1
Epichloë amarillans	AFRF00000000.1
E. aotearoae	http://www.endophyte.uky.edu/
E. baconii	http://www.endophyte.uky.edu/
E. brachyelytri	AFRB00000000.1
E. elymi	AMDJ00000000.1
E. festucae E2368	ADFL00000000.2
E. festucae Fl1	AFRX00000000.1
E. gansuensis	AFRE00000000.1
E. gansuensis var. *inebrians*	AMDK00000000.1
E. glyceriae E277	AFRG00000000.1
E. mollis	http://www.endophyte.uky.edu/
E. typhina E8	AMDI00000000.1
E. typhina subsp. *poae* E5819	AFSE00000000.1
E. typhina subsp. *poae* Ps.1	KJ502561
Photorhabdus asymbiotica Mcf1	CAQ85457
Ph. luminescens Mcf1	AAM88787
Ph. luminescens Mcf2	AAR21118
Pseudomonas chlororaphis subsp. *aureofaciens* FitD	AFD97973
Ps. protegens FitD	ABY91230
Xenorhabdus bovienii Mcf1	YP 003468304
X. nematophila Mcf1	YP 003712501
X. nematophila Mcf2	CBJ90088

Table 3.3. Estimation of the divergence rate of the basal *Epichloë* spp. relative to *Atkinsonella hypoxylon*.

Gene	Length, nt	Species	K_s	K_a	Average K_s	Divergence rate
Glutamate dehydrogenase	3,204	*E. gansuensis*	1.290	0.042	1.251	7.72×10^{-9}
		E. glyceriae	1.211	0.042		
Isoleucine tRNA synthase	3,216	*E. gansuensis*	1.236	0.048	1.243	7.67×10^{-9}
		E. gansuensis var *inebrians*	1.243	0.049		
		E. glyceriae	1.249	0.050		
MCM7	2,441	*E. gansuensis*	0.957	0.040	0.961	5.93×10^{-9}
		E. gansuensis var *inebrians*	0.943	0.040		
		E. glyceriae	0.983	0.040		
TSR1	2,487	*E. gansuensis*	1.080	0.075	1.050	6.48×10^{-9}
		E. gansuensis var *inebrians*	1.037	0.073		
		E. glyceriae	1.033	0.070		

Table 3.4. Estimation of the divergence times between the basal *Epichloë* spp. and *Claviceps purpurea.*

Gene	Species	K_s	K_a	Average K_s	Divergence time, Mya
Glutamate dehydrogenase	*E. gansuensis*	0.841	0.048	0.830	53.7
	E. glyceriae	0.818	0.049		
Isoleucine tRNA synthase	*E. gansuensis*	0.918	0.058	0.939	61.2
	E. gansuensis var *inebrians*	0.927	0.061		
	E. glyceriae	0.971	0.059		
MCM7	*E. gansuensis*	0.749	0.052	0.766	64.6
	E. gansuensis var *inebrians*	0.786	0.051		
	E. glyceriae	0.764	0.052		
TSR1	*E. gansuensis*	0.735	0.073	0.720	55.6
	E. gansuensis var *inebrians*	0.705	0.074		
	E. glyceriae	0.720	0.070		

Mean divergence time, Mya: 58.8 ± 2.5

Table 3.5. Estimation of the divergence time between *E. gansuensis* and *E. glyceriae.*

Gene	**Species**	K_s	K_a	**Average K_s**	**Divergence time, Mya**
Glutamate dehydrogenase	*E. gansuensis*	0.108	0.017	0.108	6.99
Isoleucine tRNA synthase	*E. gansuensis*	0.080	0.009	0.080	5.22
	E. gansuensis var *inebrians*	0.080	0.011		
MCM7	*E. gansuensis*	0.117	0.010	0.117	9.87
	E. gansuensis var *inebrians*	0.117	0.010		
TSR1	*E. gansuensis*	0.094	0.017	0.087	6.71
	E. gansuensis var *inebrians*	0.080	0.016		

Mean divergence time, Mya: 7.2 ± 0.97

The *Epichloë mcf-like* gene donor is likely a bacterium related to *Xenorhabdus* or *Photorhabdus*

Identification of the donor in a HGT event can often be achieved by the phylogenetic analysis of the gene involved (Acuña et al., 2012; Pauchet et al., 2013). A robust phylogenetic reconstruction of the putative laterally transferred gene and its homologs will invariably reveal atypical placements of taxa in trees of the gene vis–à–vis the organismal speciation. In other words, the gene phylogeny would be in conflict with the history of the organism. This phylogenetic incongruence is often the gold standard used to identify HGT (Keeling and Palmer, 2008). Mcf1 and Mcf2 from *Photorhabdus luminescens*, and FitD from *Pseudomonas* spp. were the best matches to the *Epichloë* Mcf-like amino acid sequences. Hence, the genes encoding these proteins were considered xenologs, (xeno = foreign) a term used to describe homologous genes

arising from HGT (Gray and Fitch, 1983). There was almost identical overall similarity numbers between the fungal *Epichloë* Mcf-like protein sequences to each of the bacterial proteins. Similar Mcf2 proteins were additionally found in *Xenorhabdus* spp., also bacterial symbionts of entomopathogenic nematodes (Hinchliffe et al., 2010) The function of these Mcf2 proteins remains to be investigated.

Maximum parsimony phylogenetic analysis of the bacterial Mcf proteins and *Epichloë* Mcf-like proteins is shown in Figure 3.4. The tree is rooted with the protein sequence of *Vibrio tubiashii* due to its considerably lower similarity to the *Epichloë* sequences compared to the *Xenorhabdus*, *Photorhabdus*, and *Pseudomonas* sequences. Additional phylogenies obtained using the neighbor joining, and branch and bound methods reflected the same relationship among taxa (Figure 3.5). Accession numbers of the sequences used are given in Table 3.2.

Surprisingly, the placement of the Mcf-like sequences from *E. gansuensis* var. *inebrians* and *E. typhina* subsp. *poae* in the phylogenies was basal to the *Xenorhabdus* spp. Mcf2 and *Photorhabdus* spp. Mcf proteins. This hindered the identification of a specific bacterial donor, and seemed to imply that the bacterial donor may be a hitherto unknown, or an already extinct bacterial species related to extant *Xenorhabdus* or *Photorhabdus* spp. In a phylogenetic comparison of many *Pseudomonas* spp., only one clade had *fitD* genes, and the authors suggested they were acquired through horizontal gene transfer (Loper et al., 2012). In the phylogenies shown in this study, *Pseudomonas* FitD sequences cluster together with *Photorhabdus* Mcf1 (Figures 3.4 and 3.5). These sequences also share an overall size similarity. Taken together, *Pseudomonas* may have acquired *fitD* laterally from a *Photorhabdus* sp. donor.

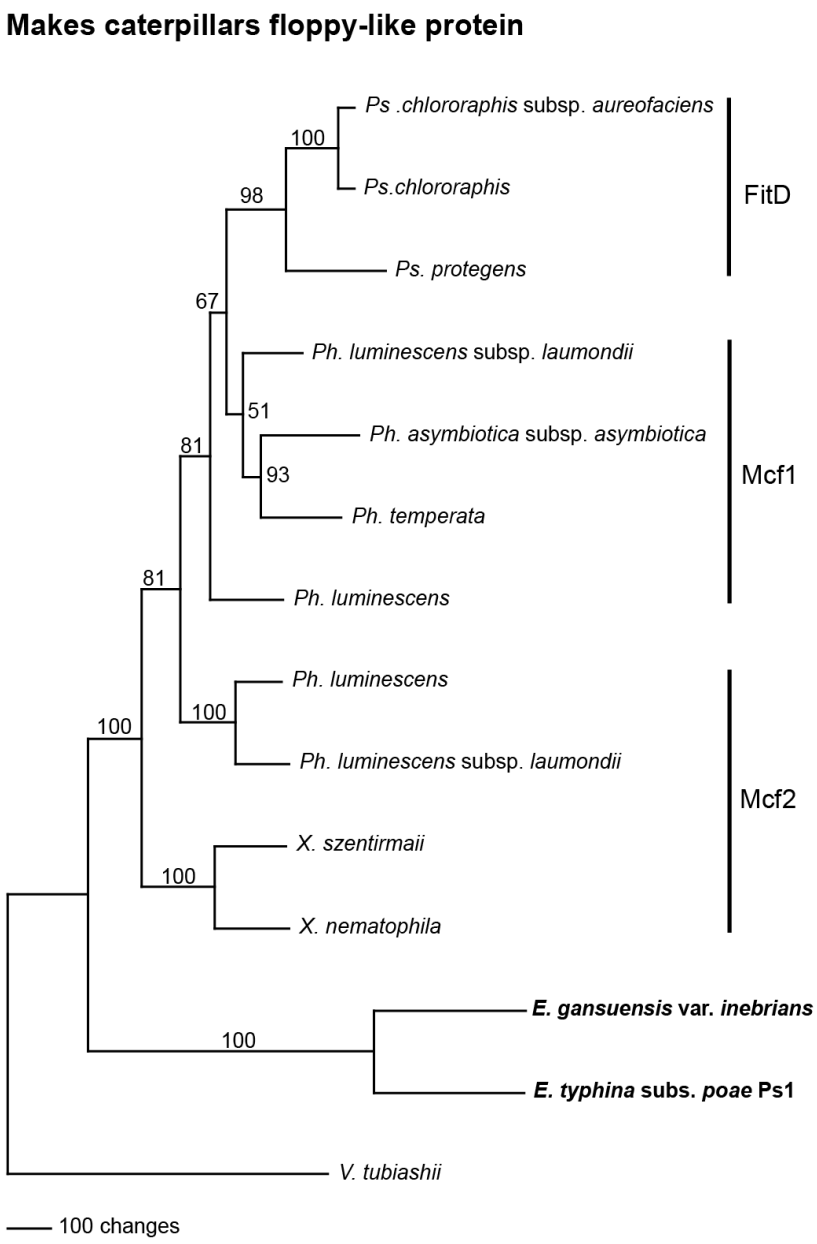


Figure 3.4. Rooted 50% majority rule maximum parsimony phylogenetic tree of bacterial Mcf proteins and the Mcf-like protein sequences from *E. gansuensis* var. *inebrians* and *E. typhina* subsp. *poae.* The *Vibrio tubiashii* sequence was designated as the outgroup for rooting the tree. The numbers at the nodes are the bootstrap percentages based on 1,000 replications. The tree was based upon 1,983 total characters, of which 391 were constant, 404 variable characters were parsimony uninformative, and 1,188 variable characters were parsimony informative.

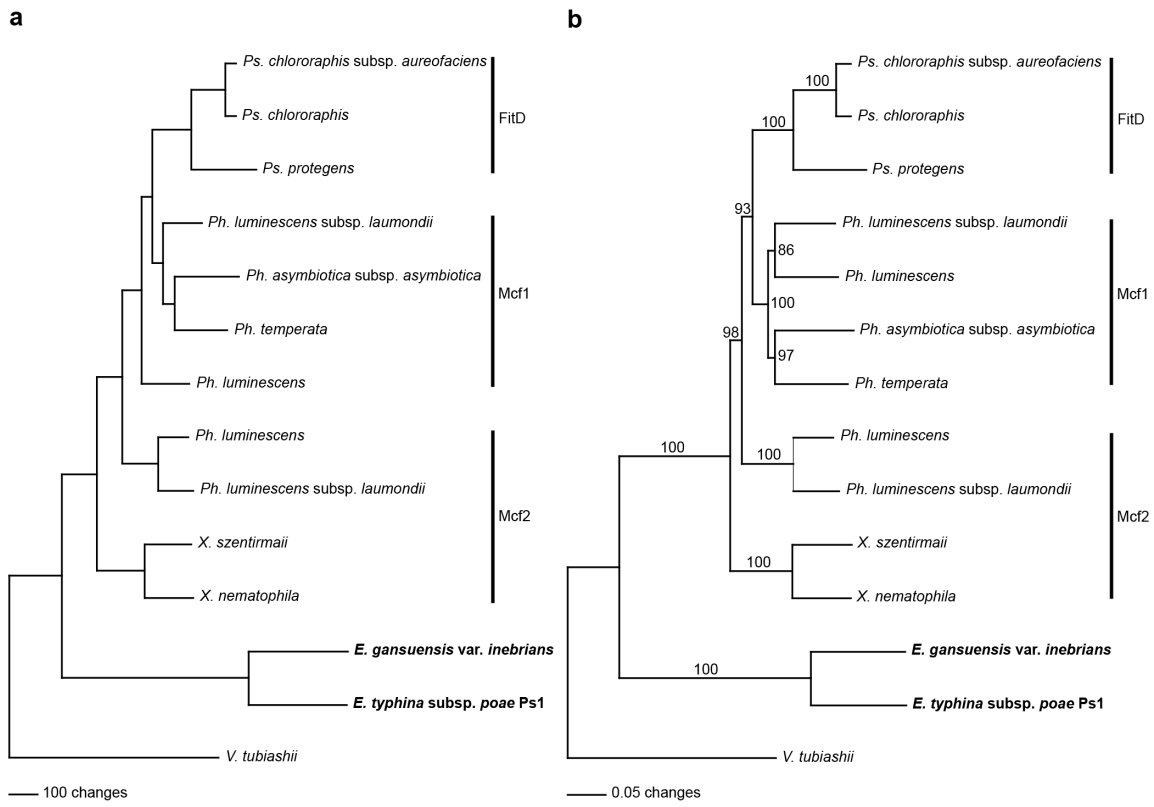


Figure 3.5. Mcf phylogenetic trees generated by the (a) branch and bound, and (b) neighbor joining methods. The *Vibrio tubiashii* sequence was designated as the outgroup for rooting the trees. The numbers at the nodes of the neighbor joining tree are the bootstrap percentages based on 1,000 replications. The trees were based upon 1,983 total characters, of which 391 were constant, 404 variable characters were parsimony uninformative, and 1,188 variable characters were parsimony informative.

Expression and activity of the *Epichloë mcf-like* gene

The discovery of the fungal *mcf-like* gene was made during a targeted screen for putative horizontally transferred genes in the *E. festucae* 2368 genome (Schardl et al., 2013). A tag similar to those bacterial genes was also recovered in our SOLiD-SAGE quantitative transcriptome expression data (Ambrose and Belanger, 2012). Since the presence of a SOLiD-SAGE tag representing the gene indicated its expression *in planta*, total cDNA of the Rose City endophyte-

infected red fescue was used to obtain the *mcf-like* transcript of the fungal isolate. In order to confirm gene expression, the *mcf-like* cDNA was PCR amplified and sequenced. The expression level of the *E. festucae* Rose City *mcf-like* cDNA was very low and required three rounds of amplification to obtain sufficient template for DNA sequencing. Further analysis of the cDNA sequences showed the presence of an additional cytosine that caused a frame-shift and an eventual early termination, relative to the other two *E. festucae* isolates (Figure 3.1a). The single nucleotide insertion was confirmed by sequencing of that region using *E. festucae* Rose City genomic DNA. This attempt revealed that although the *mcf-like* gene is expressed in *E. festucae* Rose City, a full-length protein is not produced.

Subsequent PCR amplification of the *mcf-like* gene in *E. typhina* subp. *poae* Ps1 using its host plant *P. secunda* subsp. *juncifolia* total cDNA showed that one PCR run was sufficient to obtain an appreciable amount of *E. typhina* subsp. *poae* Ps1-*mcf* cDNA for DNA sequencing (Figure 3.6a). Nucleotide sequence analysis of the amplified cDNA showed a putative full-length protein and confirmed the positions of the two annotated introns. The genomic DNA of *E. typhina* subsp. *poae* Ps1-*mcf* was also sequenced to obtain intron lengths and sequences. The gene is 5,991 bp and encodes a large 1,997 amino acid protein. The complete nucleotide sequence of *E. typhina* subsp. *poae* Ps1-*mcf* can be found as accession KJ502561 in the NCBI database.

Since *E. typhina* subsp. *poae* Ps1-*mcf* was predicted to produce an intact full-length protein encoding the toxic TcdA/TcdB pore-forming domain, an evaluation of its effect against insects was undertaken. The *E. typhina* subsp. *poae* Ps1-*mcf* cDNA was cloned into the cold-shock protein, CspA-driven pCold II expression vector and propagated in *E. coli* BL21-CodonPlus (DE3)-RIPL to induce the expression of *E. typhina* subsp. *poae* Ps1-Mcf. Functional characterization of the endophyte Mcf followed the strategy used in determining the insect toxicity activity of *Ps. protegens* FitD (Pechy-Tarr et al., 2008). Following protein induction, the

presence of a band at the expected size of 223 kD was visualized via SDS-PAGE (Figure 3.6b). The induced *E. typhina* subsp. *poae* Ps1-Mcf was found to be localized in the insoluble fraction (inclusion bodies) similar to that reported for *Ps. protegens* FitD (Pechy-Tarr et al., 2008).

Upon successful *E. typhina* subsp. *poae* Ps1-Mcf induction, the toxicity of the fungal protein against black cutworms, *Agrotis ipsilon* was tested by injecting 50 million *E. coli* cells harboring either the pCold II vector only or cells expressing the *E. typhina* subsp. *poae* Ps1-Mcf protein directly into the insect hemocoel (Daborn et al., 2002; Waterfield et al., 2003; Pechy-Tarr et al., 2008). The insect assays were done in collaboration with entomologist Dr. Albrecht Koppenhöfer, Rutgers University. The insect assays showed that injection of cells expressing *E. typhina* subsp. *poae* Ps1-Mcf resulted in death to the cutworms, while *E. coli* cells containing the pCold II vector only were harmless to the larvae. The toxic effect of the endophyte Mcf protein compared to the effect seen with the pCold II vector only against the black cutworms was statistically significant (Table 3.6; Figure 3.7). The insect assays established the threshold for black cutworms mortality as a result of *E. typhina* subsp. *poae* Ps1-Mcf toxicity since the cutworms survived injections containing a lower number of *E. coli* cells (20 million) expressing the protein.

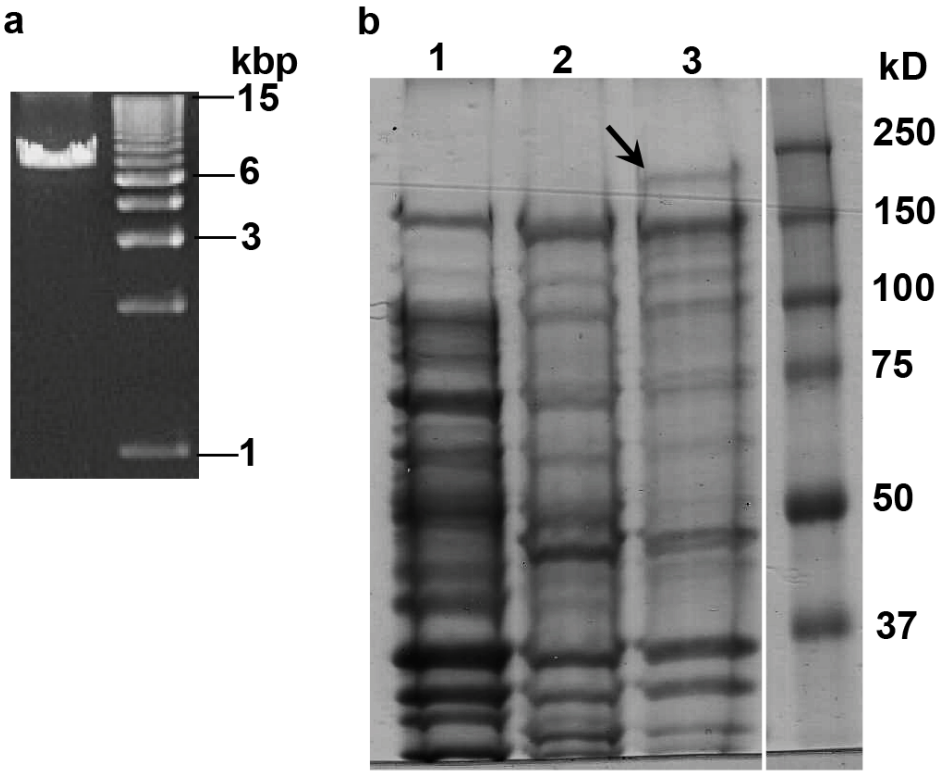


Figure 3.6. Expression of *E. typhina* subsp. *poae* Ps1-*mcf in vivo* and in *E. coli*. (**a**) PCR product of the *E. typhina* subsp. *poae* Ps1-*mcf* transcript using infected *Poa secunda* subsp. *juncifolia* plant cDNA generated from oligo(dT) as template. (**b**) SDS-PAGE analysis of insoluble *E. coli* proteins. Lane 1: *E. coli* containing the pCold II vector only subjected to induction conditions. Lane 2: *E. coli* containing the pCold II::*E. typhina* subsp. *poae* Ps1-Mcf with no induction of protein expression. Lane 3: *E. coli* containing the pCold II::*E. typhina* subsp. *poae* Ps1-Mcf subjected to induction conditions. Arrow indicates presence of the induced protein at the expected size of 223 kD.

Table 3.6. Cumulative numbers of black cutworm larvae deaths after each treatment. Insect assays were performed by injecting either sterile water, $5x10^7$ total induced *E. coli* cells containing the pCold II vector only as a negative control or the Etp-Mcf::pCold II plasmid in a total volume of 5 µL. The experiments also included larvae that were untreated.

Black cutworm mortality assay				
Experiment 1 (n = 6; *P* = 0.001)				
Treatment/Time	24 h		48 h	
Untreated	0		0	
Water	0		0	
pCold II vector only::*E. coli*	0		0	
Etp-Mcf::pCold II::*E. coli*	3		6	
Experiment 2 (n = 10; *P* = 0.029)				
Treatment/Time	24 h	48 h	72 h	120 h
Untreated	0	0	0	0
Water	0	0	0	0
pCold II vector only::*E. coli*	0	0	1	1
Etp-Mcf::pCold II::*E. coli*	1	1	5	6

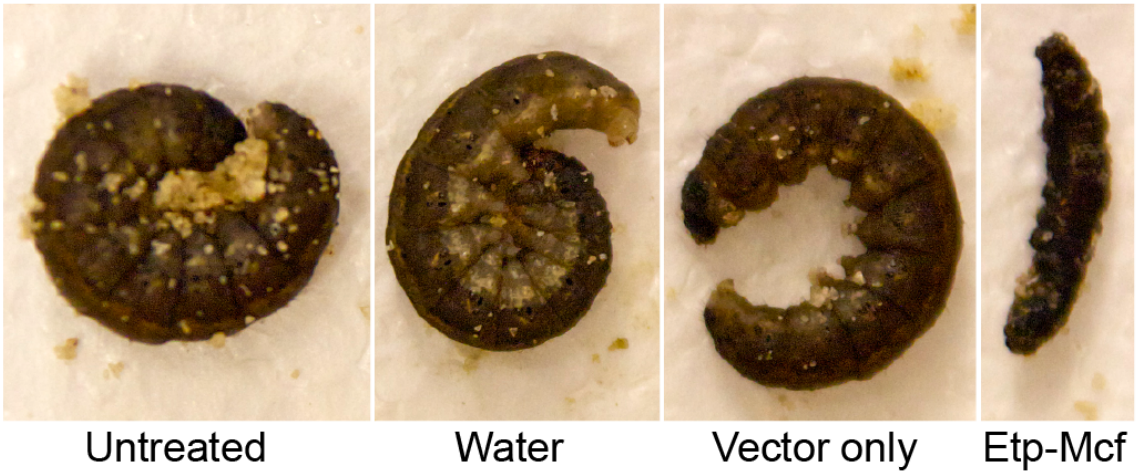


Figure 3.7. Toxicity of the *E. typhina* subsp. *poae* Ps1-Mcf protein to black cutworms. Fourth instar cutworm larvae were injected with water, 50 million *E. coli* cells containing the pCold II vector only, or 50 million *E. coli* cells expressing the *E. typhina* subsp. *poae* Ps1-Mcf protein. The live cutworms curl in response to prodding.

Discussion

The discovery and functional characterization of *mcf-like* gene in the *Epichloë* fungal endophytes of grasses that is the result of a HGT event from a bacterial donor are discussed in this chapter. The gene is present in the genomes of all 13 of the *Epichloë* spp. for which whole genome sequence data are available, as well as two additional isolates used in this study, but in no other fungal species. In 11 of the 15 *Epichloë* sequences, a long protein, containing the TcdA/TcdB pore-forming region, is predicted. In four of the *Epichloë* sequences, there was a mutation leading to an early termination codon. In three out of these four cases, multiple isolates of the same species provided examples of early termination of the *mcf-like* gene and also of a gene capable of producing a potentially functional protein. Such isolate specific breakdown of the *mcf-like* gene in *Epichloë* does not seem be an isolated phenomenon as similar breakdown of some alkaloid biosynthetic genes has been reported within the genus (Schardl et al., 2013).

Phylogenetic analysis suggests the donor bacterium may have been an as yet unknown predecessor of extant *Xenorhabdus* or *Photorabdus* spp. The mechanism underlying the HGT event is unknown, but the donor and recipient species must be in close contact at some point in time to facilitate the transfer. It is possible that the recipient *Epichloë* sp. could have come in contact with the bacterial donor either in soil or in association with a plant host. The single acquisition event of the *mcf-like* gene into the recipient *Epichloë* sp. is calculated, based on phylogenetic analyses and divergence time estimates, to have occurred between 7.2 and 58.8 Mya. The four phylogenies based on optimal fungal genes (Aguileta et al., 2008) also provide insights into the speciation of *Epichloë*. In all of the trees, there is a clear divergence of the *Epichloë* spp. into two clades, one comprised of the two *E. gansuensis* isolates and the other consisted of the rest of the *Epichloë* spp., with *E. glyceriae* at its base. The phylogenetic analyses suggest that an as yet unknown (or possibly extinct) species was the predecessor of *E. gansuensis*

and *E. glyceriae* and that the rest of the currently known *Epichloë* spp. were derived from *E. glyceriae*.

Gene acquisition via HGT that confers a new functional ability to the recipient species is important in the evolution of organisms. Several reports have expanded our knowledge in regards to these occurrences in nature. As in the case described earlier, HGT of a bacterial mannanase gene facilitated adaptation of the recipient species, the coffee borer beetle, to a new host, the mannans-rich coffee berry, thus providing a new source of nutrients for the insect (Acuña et al., 2012). In another example, Loftus et al. (2005) described 96 genes that had undergone lateral transfer into *Entamoeba histolytica*, an anaerobic parasitic protist of the human gut. Over half of those acquired genes had a bacterial origin and were involved in metabolic adaptation of the protozoan. The metabolic enzymes encoded by these laterally transferred genes were postulated to allow the major human pathogen to expand on the range of available substrates to obtain energy and breakdown sugars other than glucose. Also, recently Li et al. (2014) presented evidence that horizontal transfer of a gene encoding a unique type of photoreceptor enabled ferns to adapt to low-light environments. The hornworts, belonging to the bryophyte lineage, were the donor spp. of the chimeric neochrome that arose from the coupling of red-sensing phytochrome with blue-sensing phototropin units as one single gene. The HGT event that allowed enhanced response to light in shaded areas was considered to be advantageous for promoting the evolution and diversification of ferns.

In this study, it has been demonstrated that injection of *E. coli* expressing the *E. typhina* subsp. *poae* Ps1-*mcf* gene into black cutworms (*Agrotis ipsilon*), a pest of various crops, results in death of the insect, similar to the insecticidal activities of the bacterial *mcf1*, *mcf2,* and *fitD* genes (Daborn et al, 2002; Waterfield et al., 2002; Pechy-Tarr et al., 2008). The bacterial Mcf proteins are considered critical components of the overall insect toxicity of the bacteria. The *E. typhina*

subsp. *poae* Ps1-Mcf protein has been maintained as an active protein over millions of years since its acquisition by an ancestral *Epichloë* sp. This suggests it may also be a component, in addition to alkaloids that are well established as having insecticidal activity (Schardl et al., 2013), in the insect resistance observed in plants containing this endophyte isolate (Yue et al., 2000). Other *Epichloë* spp. have genes apparently capable of producing Mcf-like proteins that may also have activity against insects, as shown here for the *E. typhina* subsp. *poae* Ps1 isolate.

The *E. typhina* subsp. *poae* Ps1-*mcf* work described here is the first functional characterization of a toxin-encoding gene acquired through HGT by *Epichloë* with possible adaptive implications for the host plants. Understanding the role of the toxic *E. typhina* subsp. *poae* Ps1-*mcf* gene in plant genotypes infected with endophyte isolates lacking the peramine and lolitrem insecticidal alkaloids, in addition to the protein's mode of action warrants further exploration of this area. The analyses of the laterally transferred gene found in the plant's fungal endosymbiont points to an origin that is unequivocally bacterial, illuminating an intriguing tripartite interaction. Such HGTs enrich our understanding of plant-fungal symbiosis beyond bidirectional host-symbiont transfers that typically occur due to their intimate association (Richards et al., 2009). Consequently, horizontal acquisitions and maintenance of functional genes like the *E. typhina* subsp. *poae* Ps1-*mcf* from donor spp. not known to be involved in the endophyte-turf grass mutualism may have important ramifications on the macroevolution and innovation of symbiosis.

References:

1. Acuña R, Padilla BE, Flórez-Ramos CP, Rubio JD, Herrera JC, Benavides P, Lee S-J, Yeats TH, Egan AN, Doyle JJ, Rose JKC (2012) Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. Proceedings of the National Academy of Sciences USA, 109:4197–202.

2. Aguileta G, Marthey S, Chiapello H, Lebrun MH, Rodolphe F, Fournier E, Gendrault-Jacquemard A, Giraud T (2008) Assessing the performance of single-copy genes for recovering robust phylogenies. Systematic Biology, 57:613-627.

3. Alabert C, Groth A (2012) Chromatin replication and epigenome maintenance. Nature Reviews Molecular Cell Biology, 13:153-167.

4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of Molecular Biology, 215:403-410.

5. Ambrose KV, Belanger FC (2012) SOLiD-SAGE of endophyte-infected red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins. PLoS ONE, 7:e53214.

6. Blaxter M (2007) Symbiont genes in host genomes: fragments with a future? Cell Host and Microbe 2. Elsevier Inc.

7. Brunner S, Fengler K, Morgante M, Tingey S, Rfalski A (2005) Evolution of DNA sequence nonhomologies among maize inbreds. Plant Cell, 17:343-360.

8. Christin P-A, Edwards EJ, Besnard G, Boxall SF, Gregory R, Kellogg EA, Hartwell J, Osborne CP (2012) Adaptive evolution of C(4) photosynthesis through recurrent lateral gene transfer. Current Biology, 22:445–9.

9. Collins T, Azevedo-Silva J, da Costa A, Branca F, Machado R, Casal M (2013) Batch production of silk-elastin-like protein in *E. coli* BL21(DE3): key parameters for optimisation. Microbial Cell Factories, 12:21.

10. Cortez D, Forterre P, Gribaldo S (2009) A hidden reservoir of integrative elements is the major source of recently acquired foreign genes and ORFans in archaeal and bacterial genomes. Genome Biology, 10:65.

11. Da Lage J-L, Binder M, Hua-Van A, Janecek S, Casane D (2013) Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to Basidiomycetes. BMC Evolutionary Biology, 13:40.

12. Daborn PJ, Waterfield N, Silva CP, Sharma S, ffrench-Constant RH (2002) A single Photorhabdus gene *makes caterpillars floppy* (*mcf*) allows *Escherichia coli* to persist within and kill insects. Proceedings of the National Academy of Sciences USA, 99:10742-10747.

13. Dowling AJ, Daborn PJ, Waterfield NR, Wang P, Streuli CH, ffrench-Constant RH (2004) The insecticidal toxin Makes caterpillars floppy (Mcf) promotes apoptosis in mammalian

cells. Cellular Microbiology, 6:345-353.

14. Fawal N, Savelli B, Dunand C, Mathé C (2012) GECA: a fast tool for Gene Evolution and Conservation Analysis in eukaryotic protein families. Bioinformatics, 28:1398-1399.

15. Fitzpatrick DA, Logue ME, Butler G (2008) Evidence of recent interkingdom horizontal gene transfer between bacteria and *Candida parapsilosis*. BMC Evolutionary Biology, 8:181.

16. ffrench-Constant RH, Dowling A, Waterfield NR (2007) Insecticidal toxins from *Photorhabdus* bacteria and their potential use in agriculture. Toxicon, 49:436-451.

17. Frost LS, Leplae R, Summers AO, Toussaint A (2005) Mobile genetic elements: the agents of open source evolution. Nature Reviews Microbiology, 3:722-732.

18. Gelperin D, Horton L, Beckman J, Hensold J, Lemmon SK (2001) Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. RNA, 7:1268-1283.

19. Genisyuerek S, Papatheodorou P, Guttenberg G, Schubert R, Benz R, Aktories K (2011) Structural determinants for membrane insertion, pore formation and translocation of *Clostridium difficile* toxin B. Molecular Microbiology, 79:1643-1654.

20. Ginns J (2011) *Balansia hypoxylon* (Ascomycota: Clavicipitaceae) in Canada: a mycosymbiont with *Danthonia spicata* (Poaceae). North American Fungi, 6:1-6.

21. Guo AY, Zhu QH, Chen X, Luo JC (2007) GSDS: A gene structure display server. Yi Chuan, 29:1023-1026.

22. Gray GS, Fitch WM (1983) Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from *Staphylococcus aureus*. Molecular Biology and Evolution, 1:57-66.

23. Hinchliffe SJ, Hares MC, Dowling AJ, ffrench-Constant RH (2010) Insecticidal toxins from the *Photorhabdus* and *Xenorhabdus* bacteria. The Open Toxicology Journal, 3:101-118.

24. Kasuga T, White TJ, Taylor JW (2002) Estimation of nucleotide substitution rates in eurotiomycete fungi. Molecular Biology and Evolution, 19:2318-2324.

25. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nature Review Genetics, 9:605-618.

26. Korpimaki T, Kurittu J, Karp M (2003) Surprisingly fast disappearance of β-lactam selection pressure in cultivation as detected with novel biosensing approaches. Journal of Microbiological Methods, 53:37-42.

27. Kristiansson E, Fick J, Janzon A, Grabic R, Rutgersson C, Weijdegård B, Söderström H, Joakim Larsson DG (2011) Pyrosequencing of antibiotic-contaminated river sediments reveals high levels of resistance and gene transfer elements. PLoS ONE, 6:e17038.

28. Kuldau GA, Liu J-S, White JF Jr, Siegel MR, Schardl CL (1997) Molecular systematics of Clavicipitaceae supporting monophyly of genus *Epichloë* and form genus *Ephelis*. Mycologia, 89:431-441.

29. Lawrence JG, Ochman H (1997) Amelioration of bacterial genomes: rates of change and exchange. Journal of Molecular Evolution, 44:383–397.

30. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature, 227:680-685.

31. Li W-H (1997) Molecular evolution. Sinauer Associates, Sunderland, MA.

32. Li F-W, Villarreal JC, Kelly S, Rothfels CJ, Melkonian M, Frangedakis E, Ruhsam M, Sigel EM, Der JP, Pittermann J, Burge DO, Pokorny L, Larsson A, Chen T, Weststrand S, Thomas P, Carpenter E, Zhang Y, Tian Z, Chen L, Yan Z, Zhu Y, Sun X, Wang J, Stevenson DW, Crandall-Stotler BJ, Shaw AJ, Deyholos MK, Soltis DE, Graham SW, Windham MD, Langdale JA, Wong GK-S, Mathews S, Pryer KM (2014) Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proceedings of the National Academy of Sciences USA, 18:6672-6677.

33. Li X, Zhang TC, Qiao Q, Ren Z, Zhao J, Yonezawa T, Hasegawa M, Crabbe MJ, Li J, Zhong Y (2013) Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). PLoS ONE, 8:e58747.

34. Leuchtmann A, Bacon CW, Schardl CL, White JF Jr, Tadych M (2014) Nomenclatural realignment of *Neotyphodium* species with genus *Epichloë*. Mycologia, 106:202-215.

35. Loftus B, Anderson I, Davies R, Alsmark UCM, Samuelson J, Amedeo P, Roncaglia P, Berriman M, Hirt RP, Mann BJ, Nozaki T, Suh B, Pop M, Duchene M, Ackers J, Tannich E, Leippe M, Hofer M, Bruchhaus I, Willhoeft U, Bhattacharya A, Chillingworth T, Churcher C, Hance Z, Harris B, Harris D, Jagels K, Moule S, Mungall K, Ormond D, Squares R, Whitehead S, Quail MA, Rabbinowitsch E, Norbertczak H, Price P, Wang Z, Guillen N, Gilchrist C, Stroup SE, Bhattacharya S, Lohia A, Foster PG, Sicheritz-Ponten T, Weber C, Singh U, Mukherjee C, El-Sayed NM, Petri WA Jr, Clark CG, Embley TM, Barrell B, Fraser CM, Hall N (2005) The genome of the protist parasite *Entamoeba histolytica*. Nature, 7028:865-868.

36. Loper JE, Hassan KA, Mavrodi DV, Davis EW II, Lim CK, Shaffer BT, Elbourne LDH, Stockwell VO, Hartney SL, Breakwell K, Henkels MD, Tetu SG, Rangel LI, Kidarsa TA, Wilson NL, van de Mortel JE, Song C, Blumhagen R, Radune D, Hostetler JB, Brinkac LM, Durkin AS, Kluepfel DA, Wechter WP, Anderson AJ, Kim YC, Pierson LS III, Pierson EA, Lindow SE, Kobayashi DY, Raaijmakers JM, Weller DM, Thomashow LS, Allen AE, Paulsen IT (2012) Comparative genomics of plant-associated *Pseudomonas* spp.: insights into diversity and inheritance of traits involved in multitrophic interactions. PLoS Genetics, 8:e1002784.

37. Marcet-Houben M, Gabaldon T (2010) Acquisition of prokaryotic genes by fungal genomes. Trends in Genetics, 26:5-8.

38. Marthey S, Aguileta G, Rodolphe F, Gendrault A, Giraud T, Fournier E, Lopez-Villavicencio M, Gautier A, Lebrun M-H, Chiapello H (2008) FUNYBASE: a FUNgal phYlogenomic dataBASE. BMC Bioinformatics, 9:456.

39. McDonald TR, Dietrich FS, Lutzoni F (2012) Multiple horizontal gene transfers of ammonium transporters/ammonia permeases from prokaryotes to eukaryotes: toward a new functional and evolutionary classification. Molecular Biology and Evolution, 29:51-60.

40. Moy M, Li H., Sullivan R, White JF Jr, Belanger FC (2002) Endophytic fungal β-1,6-glucanase expression in the infected host grass. Plant Physiology, 130:1298-1308.

41. Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. Science, 328:624-627.

42. Moran Y, Fredman D, Szczesny P, Grynberg M, Technau U (2012) Recurrent horizontal transfer of bacterial toxin genes to eukaryotes. Molecular Biology and Evolution, 29:2223-2230.

43. Pauchet Y, Heckel DG (2013) The genome of the mustard leaf beetle encodes two active xylanases originally acquired from bacteria through horizontal gene transfer. Proceedings of the Royal Society B, 280:1763.

44. Pechy-Tarr M, Bruck DJ, Maurhofer M, Fischer E, Vogne C, Henkels MD, Donahue KM, Grunder J, Loper JE, Keel C (2008) Molecular analysis of a novel gene cluster encoding an insect toxin in plant-associated strains of *Pseudomonas fluorescens*. Environmental Microbiology, 10:2368-2386.

45. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. Nucleic Acids Research, 40:290–301.

46. Raja HA, Schoch CL, Hustad VP, Shearer CA, Miller AN (2011) Testing the phylogenetic utility of MCM7 in the Ascomycota. MycoKeys, 1:63-94.

47. Richards TA, Soanes DM, Jones MD, Vasieva O, Leonard G, Paszkiewicz K, Foster PG, Hall N, Talbot NJ (2011) Horizontal gene transfer facilitated the evolution of plant parasitic mechanisms in the oomycetes. Proceedings of the National Academy of Sciences USA, 108:15258-15263.

48. Ramette A, Frapolli M, Fischer-Le Saux M, Gruffaz C, Meyer J-M, Défago G, Sutra L, Moënne-Loccoz Y (2011) *Pseudomonas protegens* sp. nov., widespread plant-protecting bacteria producing the biocontrol compounds 2,4-diacetylphloroglucinol and pyoluteorin. Systematic and Applied Microbiology, 34:180-188.

49. Richards TA, Soanes DM, Foster, PG, Leonard G, Thornton CR, Talbot NJ (2009) Phylogenomic analysis demonstrates a pattern of rare and ancient horizontal gene transfer between plants and fungi. The Plant Cell Online, 7:1897-1911.

50. Ruffner B, Péchy-Tarr M, Ryffel F, Hoegger P, Obrist C, Rindlisbacher A, Keel C, Maurhofer M (2013) Oral insecticidal activity of plant-associated pseudomonads.

Environmental Microbiology, 15:751-763.

51. Sage RF (2004) The evolution of C4 photosynthesis. New Phytologist, 161:341–370.

52. Saha DC, Johnson-Cicalese JM, Halinsky PM, van Heemstra MI, Funk CR (1987) Occurrence and significance of endophytic fungi in the fine fescues. Plant Disease, 71:1021–1024.

53. Schardl CL (2001) *Epichloë festucae* and related mutualistic symbionts of grasses. Fungal Genetics and Biology, 33:69–82.

54. Schardl CL, Grossman RB, Nagabhyru P, Faulkner JR, Mallik UP (2007) Loline alkaloids: currencies of mutualism. Phytochemistry, 68:980–996.

55. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk JW, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J, Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach J, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z. (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the *Clavicipitaceae* reveals dynamics of alkaloid loci. PLoS Genetics, 9(2):e1003323.

56. Schmitt I, Crespo A, Divakar PK, Fankhauser JD, Herman-Sackett E, Kalb K, Nelsen MP, Nelson NA, Rivas-Plata E, Shimp AD, Widhelm T, Lumbsch HT (2009) New primers for promising single-copy genes in fungal phylogenetics and systematics. Persoonia, 23:35-40.

57. Senchina DS, Alvarez 1, Cronn RC, Liu B, Rong J, Noyes JD, Paterson AH, Wing RA, Wilkins TA, Wendel JF (2003) Rate variation among nuclear genes and the age of polyploidy in *Gossypium*. Molecular Biology and Evolution, 20:633-643.

58. Spiering MJ, Moon CD, Wilkinson HH, Schardl CL (2005) Gene clusters for insecticidal loline alkaloids in the grass-endophytic fungus *Neotyphodium uncinatum*. Genetics, 169:1403–1414.

59. Sung G-H, Poinar GO Jr, Spatafora JW (2008) The oldest fossil evidence of animal parasitism by fungi supports a Cretaceous diversification of fungal-arthropod symbioses. Molecular Phylogenetics and Evolution, 49:495-502.

60. Syvanen M (1984) The evolutionary implications of mobile genetic elements. Annual Review of Genetics, 18:271-293.

61. Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland.

62. Tadych M, Ambrose KV, Bergen MS, Belanger FC, White JF Jr (2012) Taxonomic placement of *Epichloë poae* sp. nov. and horizontal dissemination to seedlings via conidia.

Fungal Diversity, 54:117-131.

63. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution, 28:2731-2739.

64. Tanaka A, Tapper BA, Popay A, Parker EJ, Scott B (2005) A symbiosis expressed non-ribosomal peptide synthetase from a mutualistic fungal endophyte of perennial ryegrass confers protection to the symbiotum from insect herbivory. Molecular Microbiology, 57:1036–1050.

65. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research, 25:4876-4882.

66. Tretter ED, Johnson EM, Wang Y, Kandel P, White MM (2013) Examining new phylogenetic markers to uncover the evolutionary history of early-diverging fungi: comparing MCM7, TSR1, and rRNA genes for single- and multi-gene analyses of Kickxellomycotina. Persoonia, 30:106-125.

67. Waterfield NR, Daborn PJ, Dowling AJ, Yang G, Hares M, ffrench‐Constant RH (2003) The insecticidal toxin Makes caterpillars floppy 2 (Mcf2) shows similarity to HrmA, an avirulence protein from a plant pathogen. FEMS Microbiology Letters, 229:265-270.

68. Xi Z, Bradley RK, Wurdack KJ, Wong KM, Sugumaran M, Bomblies K, Rest JS, Davis CC (2012) Horizontal transfer of expressed genes in a parasitic flowering plant. BMC Genomics, 13:227. doi: 10.1186/1471-2164-13-227.

69. Xu J-H. Messing J (2008) Diverged copies of the seed regulatory *Opaque-2* gene by a segmental duplication in the progenitor genome of rice, sorghum, and maize. Molecular Plant, 1:760-769.

70. Yoshida S, Maruyama S, Nozaki H, Shirasu K (2010) Horizontal gene transfer by the parasitic plant *Striga hermonthica*. Science, 328:1128.

71. Yue Q, Gianfagna TJ, Meyer WA (2000) Alkaloid production and chinch bug resistance in endophyte-inoculated Chewings and strong creeping red fescues. Journal of Chemical Ecology, 26:279–292.

72. Zhang D, Qi J, Yue J, Huang J, Sun T, Li S, Wen J, Hettenhausen C, Wu J, Wang L, Zhuang H, Wu J, Sun G (2014) Root parasitic plant *Orobanche aegyptiaca* and shoot parasitic plant *Cuscuta australis* obtained *Brassicaceae*-specific strictosidine synthase-like genes by horizontal gene transfer. BMC Plant Biology, 14:19 doi:10.1186/1471-2229-14-19.

Chapter 4

Functional characterization of salicylate hydroxylase from *Epichloë festucae* Rose City, a fungal endosymbiont of red fescue

Chapter 4

Functional characterization of salicylate hydroxylase from *Epichloë festucae* Rose City, a fungal endosymbiont of red fescue

Introduction

Salicylic acid is one of the most important components of plant defence and immunity (Durrant and Dong, 2004) and is involved in systemic acquired resistance, SAR (Métraux et al., 1990; Raskin, 1992). The enzyme salicylate hydroxylase catalyzes the hydroxylation of salicylic acid to catechol which does not induce SAR in plants (Figure 4.1; Yamamoto et al., 1965). In an important discovery, Gaffney et al. (1993) demonstrated that transgenic tobacco plants expressing NahG, the salicylate hydroxylase from *Pseudomonas putida* PpG7, had low levels of salicylic acid and compromised ability to induce SAR against the pathogen tobacco mosaic virus (TMV), thus providing direct evidence that salicylic acid is required to induce SAR. Degradation of salicylic acid circumvented SAR induction, and rendered the transgenic tobacco plants incapacitated to respond to the TMV onslaught, since the plant hormone is an important signaling molecule to initiate defense mechanisms to counter pathogenesis.

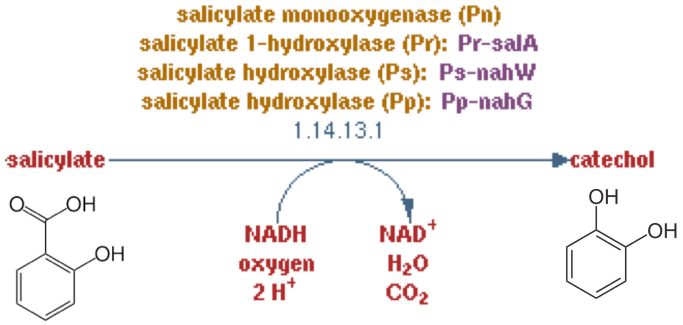


Figure 4.1. Reaction catalyzed by salicylate hydroxylase (E.C.1.14.12.1). Examples of well-known salicylate hydroxylase-encoding genes from *Pseudomonas nitroreducens*, Pn (Kohler et al., 1993), *P. reinekei*, Pr-salA (Camara et al., 2007), *P. stutzeri* AN10, Ps-nahW (Bosch et al.,

1999), and *P. putida* PpG7, Pp-nahG (Yamamoto et al., 1965; You et al., 1990) are shown in the diagram (Caspi et al., 2008).

There are few reports on fungal salicylate hydroxylases as the overwhelming majority of studies have focused on the enzyme's role in bacteria such as *Pseudomonas* spp. to biodegrade environmentally-polluting polyaromatic hydrocarbons such as naphthalene and phenanthrene (Seo et al., 2009). Fungal salicylate hydroxylase enzymatic activities have been studied in the soil yeast *Trichosporon cutaneum* (Szu and Dagley, 1984), *Aspergillus nidulans* (Graminha et al., 2004), *Fusarium* spp. (Dodge and Wackett, 2005), and *T. moniliiforme* (Iwasaki et al., 2010). In a recent study that involved the smut fungus *U. maydis*, three putative salicylate hydroxylase genes were identified in the pathogen's genome to have similarity to *Pseudomonas* spp. NahG (Rabe et al., 2013). One of the three candidates exhibited the ability to utilize salicylic acid as a substrate, and hence was considered to be an active enzyme that was subsequently named Shy1 (Salicylate hydroxylase 1). The role played by Shy1 in the pathogenesis of maize was duly investigated by quantitative transcript expression, and Shy1-knockout mutants experiments. To date, Shy1 is the only heterologously expressed, and functionally characterized fungal salicylate hydroxylase.

In the SOLiD-SAGE quantitative transcriptome study of *Epichloë festucae* RC infected-*Festuca rubra* (red fescue*)* vis-à-vis endophyte-free *F. rubra* (Ambrose and Belanger, 2012), a gene whose best match was to the *Pseudomonas* NahG salicylate hydroxylase protein was one of the highly expressed transcripts *in planta.* The rationale for the current study was based on the possibility that the fungal salicylate hydroxylase is a factor in the endophyte's ability to evade the host plant's defense mechanisms by degradation of salicylic acid. It is well known that plants infected with the fungal endophyte do not exhibit hypersensitive responses commonly associated with pathogenesis (Clay and Schardl, 2003).

The use of salicylate hydroxylase to alter plant defense mechanisms is a strategy employed by pathogens (Kunkel and Brooks, 2002), and in several cases the suppression of defence was separate from the enzyme's role in salicylic acid degradation (Heck et al., 2003; Van Wees and Glazebrook, 2003; Morse et al., 2007). Interesting recent developments have pointed out that beneficial microbes also have the capacity to modify plant hormone signaling pathways using symbiotic effectors so that the host plant immunity is altered and a prolonged mutualistic relationship is enabled (Pieterse et al., 2012). Therefore the *E. festucae* RC fungal salicylate hydroxylase was selected to be functionally characterized as the first step to understand the role of the enzyme in a symbiotic relationship such as the interaction seen between *E. festucae-F. rubra*.

In this chapter, the expression, purification and functional characterization of the salicylate hydroxylase from *E. festucae* Rose City, the fungal endosymbiont of red fescues is described. The protein has similarity to the classic *P. putida* PpG7 NahG (You et al., 1990) with 36% identity, and is an active salicylate hydroxylase.

Materials and methods

Plant and fungal samples

E. festucae-infected *Festuca rubra* plants (red fescue; S1139 Rose City genotype) were grown by clonal propagation and maintained in 15 cm pots in the greenhouse. The *E. festucae* Rose City (*E. festucae* RC) fungus was isolated from its host plant by plating surface-sterilized leaf sheath tissue on potato dextrose agar (Difco Laboratories, Detroit, MI).

DNA and RNA isolation

Genomic DNA of *E. festucae* RC was extracted from culture grown in potato dextrose broth for 14 days on a shaker (175 rpm) at room temperature (23-25°C). Fungal DNA was isolated using

phenol-chloroform as previously described (Ambrose and Belanger, 2012). One gram of endophyte mycelial cells from culture grown for 14 days or the innermost leaf sheath tissues of *E. festucae* RC-infected *F. rubra* plants were used to isolate RNA (Ambrose and Belanger, 2012). Briefly, the samples were ground to a fine powder with liquid nitrogen and resuspended in 10 mL Tri-Reagent (Sigma-Aldrich, St. Louis, MO). Debris was removed by centrifugation and supernatant was extracted twice with chloroform. RNA in the aqueous layer was precipitated with isopropanol, and the RNA pellet was washed once with ethanol and dissolved in water. Nucleic acid concentration was measured using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA).

Amplification and sequencing of *E. festucae* RC salicylate hydroxylase gene and cDNA

First-strand cDNA of 5 µg *E. festucae* RC fungal or *E. festucae* RC-infected *Festuca rubra* plant total RNA was synthesized from 500 ng of oligo$(dT)_{18}$ primer by using SuperScript™ III Reverse Transcriptase (Life Technologies, Carlsbad, CA) according to the manufacturer's instructions. PCR was performed in 50 µL reactions with either 1.0 µg of fungal genomic DNA, 1 µL *E. festucae RC fungal* cDNA or 1 µL *E. festucae RC*-infected *Festuca rubra* plant cDNA, 1X PrimeSTAR Max Premix (Clontech Laboratories, Mountain View, CA), and 0.3 µM of each forward and reverse primer (Integrated DNA Technologies, Coralville, IA). PCR was done in a GeneAmp 9700 thermocycler (Applied Biosystems, Inc., Foster City, CA) with 35 cycles of denaturation at 98°C for 10 s, followed by 30 s annealing at 55°C, and 30 s extension at 72°C. Forward and reverse primers were designed based on the *E. festucae* 2368 genome sequence (Schardl et al., 2013). The PCR products were visualized on a 1% TBE agarose gel. Each PCR product was sequenced directly (Genewiz, Inc., South Plainfield, NJ) after purification using 0.5X Agencourt AMPure XP (Beckman Coulter, Brea CA) to remove any fragments under 1 kb. For each sequencing reaction, 40 ng of purified PCR product in 10 µL was treated with 2 µL ExoSAP-IT (USB Corp., Cleveland, OH) to remove unincorporated primers and excess dNTPs.

The ExoSAP-IT reaction was performed at 37°C for 15 min followed by heating at 80°C for 15 min to inactivate the enzymes. Sequencing was done in both directions. The salicylate hydroxylase gene in *E. festucae* RC isolate was also amplified and sequenced to confirm intron lengths and positions.

Expression of the *E. festucae* Rose City salicylate hydroxylase gene in *Escherichia coli*

To test the activity of the *E. festucae* RC salicylate hydroxylase enzyme, its cDNA was cloned into the expression vector pET21b (Novagen, Billerica, MA). The *E. festucae* RC cDNA was amplified by PCR as described above with oligonucleotides that introduced a *Nde*I site at the 5' end and a *Xho*I site at the 3' end. The sequences of the oligonucleotides are: forward 5'- GATATACATATGGCGACCAAGAAAGACCA-3' and reverse 5'-TCAACTCTCGAGACCAACGGCACGCCACCT -3'. Restriction enzyme sequences are underlined. Two hundred ng of the purified PCR product was digested with the restriction enzyme *Nde*I (TaKaRa Bio, Shiga, Japan), purified with 0.5X Agencourt AMPure XP to remove any fragments under 1.0 kb, digested with *Xho*I, and again purified with 0.5X AMPure XP. The expression vector pET21b was similarly digested with *Nde*I and *Xho*I. Overnight ligation of the digested *E. festucae* RC salicylate hydroxylase PCR fragment and pET21b plasmid was done at a 2:1 insert: vector molar ratio using T4 DNA ligase (New England Biolabs, Ipswich, MA).

Two µL of the ligation product containing 4.3 ng was used to transform 20 µL *E. coli* DH5α electroporation-competent cells. The transformed cells were incubated in SOC medium for 1 hour at 37°C with shaking, followed by overnight growth of cells on LB medium supplemented with 100 µg/mL ampicillin. Transformed bacterial colonies were screened for recombinant plasmids containing the *E. festucae* RC salicylate hydroxylase gene insert by using PCR as described above. Selected clones were grown overnight in 5 mL LB supplemented with 100 µg/mL ampicillin at 37°C on a shaker (200 rpm), followed by plasmid purification using

QIAprep Spin Miniprep Kit (Qiagen, Valencia, CA). DNA sequencing of the plasmids were done as described above. A plasmid containing the correct sequence was transformed into two *E. coli* competent cells strains, these were BL21-CodonPlus (DE)-RIPL (Agilent Technologies, Santa Clara, CA) and ArcticExpress RP (DE) (Clontech Laboratories, Mountain View, CA). Transformants containing pET21b vector only were generated for use as negative control

To confirm expression of the cloned salicylate hydroxylase gene, *E. coli* BL21-CodonPlus (DE3)-RIPL transformant cells containing the *E. festucae* RC salicylate hydroxylase::pET21b recombinant plasmid were grown overnight in LB supplemented with 100 µg/mL ampicillin at 37°C on a shaker (200 rpm). Cells were streaked onto LB medium containing 100 µg/mL ampicillin, 2.5 mM salicylic acid and 0.01 mM IPTG followed by incubation at 37°C. *E. coli* cells carrying the pET21b vector only was treated the same and used as negative control. Plates were examined for accumulation of dark brown oxidation products (Cane and Williams, 1982).

Purification of the *E. festucae* Rose City salicylate hydroxylase protein in *E. coli*

For expression of the *E. festucae* RC salicylate hydroxylase protein in BL21-CodonPlus (DE3)-RIPL, transformant cells were grown overnight in LB and 100 µg/mL ampicillin at 37°C on a shaker (200 rpm). A 1:20 dilution of the overnight culture was added to fresh LB supplemented with 100 µg/mL amplicillin and incubated at 37°C on a shaker (200 rpm) until OD 0.5-0.6. To induce protein expression, isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to 0. 01 mM to the culture and then incubated at 20°C with rotational shaking for 4 hrs. *E. coli* cells expressing the recombinant protein were pelleted and lysed using 1X FastBreak Cell Lysis Reagent (Promega, Madison, WI), 0.1 mg/mL DNaseI (Sigma-Aldrich, St. Louis, MO), 200 µg/mL lysozyme (Sigma-Aldrich, St. Louis, MO) and 1X Protease Inhibitor Cocktail (Sigma-Aldrich, St. Louis, MO). Following lysis, samples were centrifuged at 10,000 RPM for 15 min to allow phase separation, and the supernatant removed. The soluble phase was treated with MagneHis Protein

Purification System (Promega, Madison, WI) to bind and purify the 6X Histidine-tagged recombinant protein following manufacturer's protocol.

E. festucae RC salicylate hydroxylase protein expression in ArcticExpress RP (DE) cells which was done to improve protein solubility. A 0.5 liter bacterial culture was grown overnight at 30°C in LB plus 100 µg/mL carbenicillin and 20 µg/mL gentamycin on a shaker (200 rpm). A 1:20 dilution of the overnight culture was added to fresh LB and 100 µg/mL carbenicillin and grown at 30°C on a shaker (200 rpm) until OD 0.5-0.6. To induce protein expression, IPTG was added to a final concentration of 0.01 mM. The culture was grown for 21 hrs at 10 °C with rotational shaking. ArcticExpress RP (DE) bacterial cells expressing the salicylate hydroxylase protein were pelleted and lysed using xTractor Cell Lysis Buffer (Clontech Laboratories, Mountain View, CA), 0.1 mg/mL DNaseI (Sigma-Aldrich, St. Louis, MO), 200 µg/mL lysozyme (Sigma-Aldrich, St. Louis, MO) and 1X Protease Inhibitor Cocktail (Sigma-Aldrich, St. Louis, MO). Following lysis, samples were centrifuged at 10,000 RPM for 30 min to allow phase separation, and the supernatant removed. The soluble phase was treated with TALON Metal Affinity Resin (Clontech Laboratories, Mountain View, CA) to bind and purify the polyHis-tagged protein of interest following manufacturer's protocol. Imidazole was removed from the purified protein sample using a disposable PD-10 desalting column (GE Healthcare Life Sciences, Pittsburgh, PA) equilibrated with 33 mM potassium phosphate buffer, pH 7. The sample was concentrated using an Amicon Ultra 4 mL centrifugal filter with a 10 kDalton molecular weight cut-off (EMD Millipore Corporation, Billerica, MA). The concentrated purified protein was quantified using Bio-Rad Protein Assay Dye Reagent Concentrate (Bio-Rad Laboratories, Hercules, CA) based on the standard Bradford dye-binding method (Bradford, 1976).

For sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) analysis, 1 mL culture of *E. coli* ArcticExpress RP (DE) expressing the recombinant protein was purified using

either MagneHis Protein Purification System or TALON Metal Affinity Resin procedure as described above. SDS sample buffer (Laemmli, 1970) was added to the samples to a final concentration of 1X, boiled for 5 minutes, and subjected to electrophoresis on a 10% gel. Gels were stained with Coommassie Blue to visualize protein bands.

For SDS-PAGE gel imaging, gels were first destained, and then dried using a modified protocol from Moghaddam and Reinton (2008). Gels were completely hydrated in a cocktail solution that consisted of 30% water, 50% methanol, and 20% polyethylene glycol 400 (PEG-400). Cellophane sheets, first soaked in water, were used to sandwich the gels in a gel-drying frame. The gels were dried at room temperature overnight. Scanned images of the SDS-PAGE gel were made using a HP Officejet Pro 8600 Plus All-in-One Printer (Hewlett-Packard, Palo Alto, CA).

Enzyme activity assays of *E. festucae* Rose City salicylate hydroxylase

Half a microgram of purified salicylate hydroxylase was used per reaction to measure catalytic activity. Activity was measured as a decrease of NADH absorbance at 340 nm using a UV160-U UV-VIS spectrophotometer (Shimadzu, Kyoto, Japan) plotted against time. Each reaction contained 150 µM of salicylic acid or 2-5-dihydroxybenzoic acid, 120 µM NADH (Sigma-Aldrich, St. Louis, MO), 10 µM FAD (Sigma-Aldrich, St. Louis, MO), and 33 mM potassium phosphate buffer, pH 7 in a final volume of 1 mL. Enzyme activity was measured as the oxidation of NADH per min measured at 25°C. The reactions were assayed in triplicate. Specific activities were expressed as nmol NADH per min per mg protein.

Growth assays of *E. festucae* Rose City fungal endophyte on salicylate minimal medium

Growth of the fungal endophyte was analysed on Yeast Nitrogen Base (Sigma-Aldrich, St. Louis, MO) supplemented with nitrogen and either 2% glucose, 10 mM salicylic acid, or without any

carbon source. Growth of the fungal endophyte was monitored to determine whether salicylic acid could be utilized as a nutrient.

Nucleic and amino acid sequence alignments

Alignments of sequences obtained from the National Center for Biotechnology Information (NCBI) or this study were performed using Clustal Omega hosted by the European Bioinformatics Institute (EMBL-EBI) (Goujon et al., 2010; Sievers et al., 2011).

Schematic representation of salicylate hydroxylase gene structure

Schematic representation of the *E. festucae* Rose City salicylate hydroxylase gene and alternatively spliced transcript variants was done using Gene Structure Display Server (Guo et al., 2007; http://gsds.cbi.pku.edu.cn/index.php?input=site).

Results

PCR of *E. festucae* Rose City salicylate hydroxylase coding sequence

Salicylate hydroxylase was one of the abundant fungal transcripts recovered in the SOLiD-SAGE transcriptome study (Ambrose and Belanger, 2012). Since expression of salicylate hydroxylase could be a factor in the reported lack of a hypersensitive response in the symbiotic association, the enzyme was chosen for further characterization.

Successful PCR amplification of the salicylate hydroxylase coding sequence from cDNA proved to be challenging initially, although the amplification of the 1,642 bp gene using genomic DNA was easily obtained. Five different DNA polymerases were tested in an attempt to obtain the correct amplified product using oligo(dT)-generated cDNA from total RNA of either the fungal endophyte or *E. festucae* RC-infected red fescues. These DNA polymerases were (i) AmpliTaq

Gold (Life Technologies, Carlsbad, CA), (ii) Elongase (Life Technologies, Carlsbad, CA), (iii) GoTaq (Promega, Madison, WI), (iv) Phusion Hot Start II High Fidelity DNA Polymerase (Thermo Fisher Scientific, Pittsburgh, PA), and (v) Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA). Owing to unknown reasons none of these five DNA polymerase were able to PCR amplify the coding sequence of the salicylate hydroxylase gene.

Robust amplification of the fungal salicylate hydroxylase coding sequence from cDNA was finally obtained using PrimeSTAR Max Premix (Clontech Laboratories). Gel analysis of the PCR revealed two bands indicating two potential transcripts for the gene. The larger PCR product was sized as approximately 1.5 kb, and a second smaller one was estimated to be around 1.3 kb (Figure 4.2). These two bands were present in both *E. festucae* RC fungal and *E. festucae* RC-infected *Festuca rubra* plant-derived samples.

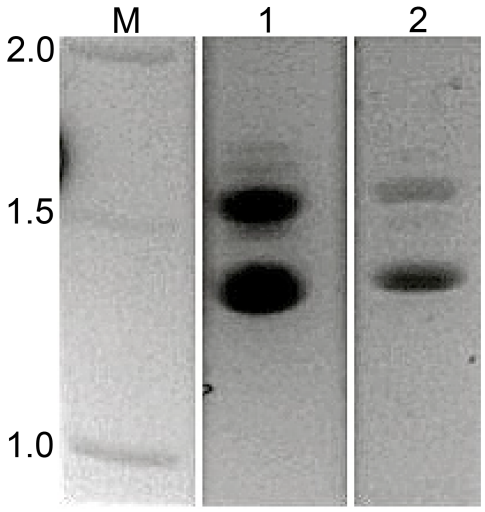


Figure 4.2. Expression of *E. festucae* Rose City salicylate hydroxylase *in vivo.* PCR product of the *E. festucae* RC salicylate hydroxylase transcript using oligo(dT)-generated cDNA reverse-transcribed from total RNA of (**1**) *E. festucae* RC fungal endophyte, and (**2**) endophyte infected-*F. rubra* plant. Standard DNA markers are indicated by M.

DNA sequencing analysis of the larger cDNA PCR product revealed that the entire 199 bp of the first intron of the gene was retained in the coding sequence, indicating an alternatively spliced transcript variant (designed Transcript variant 1 in Figure 4.3). Further investigation of the DNA sequence indicated that it was not expected to encode a fully translated salicylate hydroxylase protein due to early termination of the coding sequence at position 990. The second PCR product was also sequenced and found to be 1,296 bp. The sequence of this transcript translated into a 431 amino acid protein, and was determined to be the putative functional salicylate hydroxylase (designated Functional transcript in Figure 4.3).

The coding sequence of the putative functional transcript was subsequently cloned into pET21b vector to investigate the expression of the gene in *E. coli*. Bacterial clones hosting the salicylate hydroxylase::pET21b recombinant plasmid were screened using DNA sequencing to select for clones containing the correct 1,296 bp salicylate hydroxylase insert. The presence of a second alternatively spliced transcript of the *E. festucae* RC gene was detected at this stage (designated Transcript variant 2 in Figure 4.3). This alternately spliced form of the gene included an extra 4 bp of the 199 bp first intron that resulted in a premature termination site at position 795 of the coding sequence. Hence this transcript variant was also not expected to translate into a functional protein.

Expressed sequence tags (ESTs) resembling only the larger variant of salicylate hydroxylase transcript described here were found in the *E. festucae* endophyte isolate of meadow fescue (*Festuca pratensis*) in NCBI (Table 4.1), revealing that alternatively spliced forms of this gene were not unique to the *E. festucae* RC isolate. No ESTs of the functional transcript were recovered for any *Epichloë* isolate in NCBI. Nucleic acid alignment of *E. festucae* RC gene, and the functional form of the transcript in addition to ESTs of the *F. pratensis* isolate indicated that all five ESTs contained parts of the first intron, albeit in various lengths (Figure 4.4). The ESTs

did not always conform to the typical splice donor/acceptor consensus sequences, GT and AG, respectively. Two of the five ESTs contain the correct splice junction for the second intron, while the rest have a larger intron compared to the experimentally determined 147 bp. The observation of variations in transcript splicing of the salicylate hydroxylase gene in *E. festucae* raises questions as to the prevalence and role of such incidences in the fungal endosymbiont. Interestingly, an alternatively spliced transcript was also observed for *EfAFP*, the second most abundant *E. festucae* RC transcript expressed *in planta*, that putatively encodes an antifungal protein similar to the *Penicillium chrysogenum* antifungal protein PAF (Ambrose and Belanger, 2012).

Table 4.1. NCBI accessions of ESTs containing *E. festucae* salicylate hydroxylase. All sequences were obtained from the fungal endophyte isolate of meadow fescue (*F. pratensis*) and seemed to be alternatively spliced variants of the gene.

NCBI accession	EST length (nt)
GO805203	754
GO805370	703
GO814712	725
GO814910	730
GO851910	735

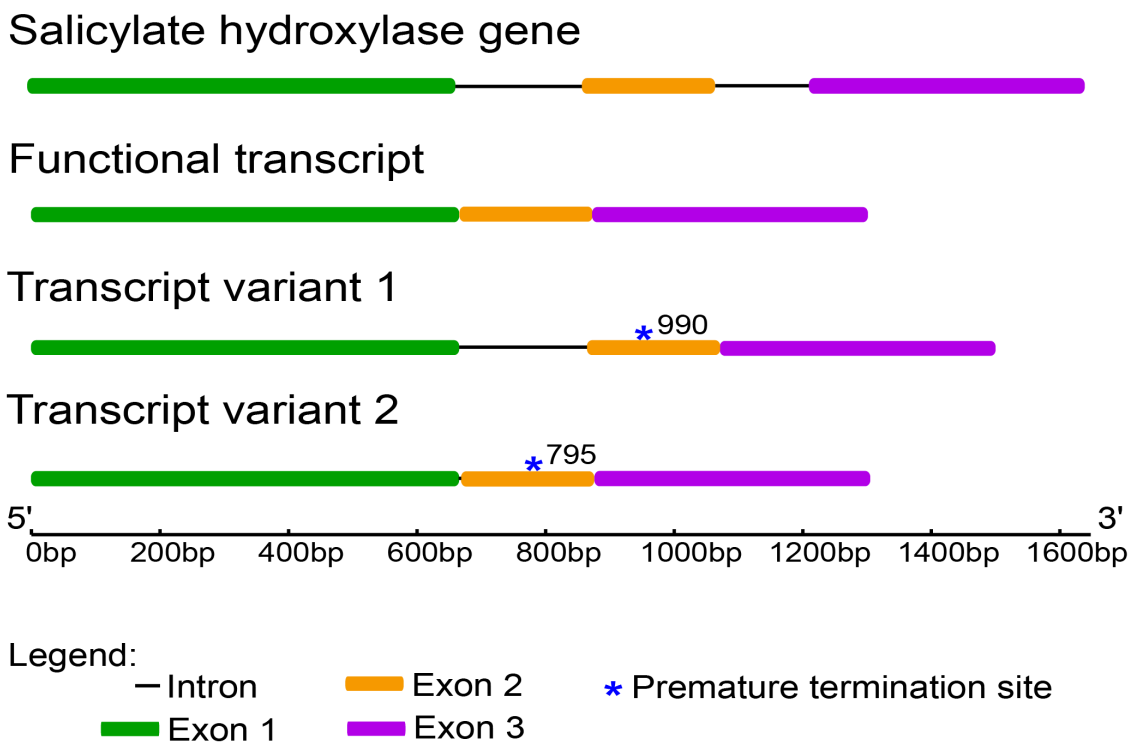


Figure 4.3. Gene structure of the *E. festucae* Rose City salicylate hydroxylase gene, functional transcript, and alternatively spliced transcript variants. The exons are indicated by boxes and the introns by lines. The position of the premature termination site in the alternatively spliced transcript variants is shown by asterisks.

Figure 4.4. Nucleic acid alignment of the experimentally determined *E. festucae* Rose City salicylate hydroxylase gene and functional transcript sequences in this study, and the five ESTs of transcript variants in the *E. festucae* isolate of *F. pratensis* obtained from NCBI (accession numbers are displayed). Exon sequences are highlighted in color.

```
Gene          ATGGCGACCAAGAAAGACCATGTTTTCGCCATAATCGGCGGCGGCATCGCTGGTCTCACG 60
Transcript    ATGGCGACCAAGAAAGACCATGTTTTCGCCATAATCGGCGGCGGCATCGCTGGTCTCACG 60
GO851910      ------------------------------------------------------------ 0
GO814712      ------------------------------------------------------------ 0
GO814910      ------------------------------------------------------------ 0
GO805203      ------------------------------------------------------------ 0
GO805370      ------------------------------------------------------------ 0

Gene          CTCGCGGTTGCCCTTCACCATCGCGGACTCAGAATCAAGATATTCGAACAAGCAGGGCAG 120
Transcript    CTCGCGGTTGCCCTTCACCATCGCGGACTCAGAATCAAGATATTCGAACAAGCAGGGCAG 120
GO851910      ------------------------------------------------------------ 0
GO814712      ------------------------------------------------------------ 0
GO814910      ------------------------------------------------------------ 0
GO805203      ------------------------------------------------------------ 0
GO805370      ------------------------------------------------------------ 0

Gene          ATGCAAGAAATCGGCGCCGGGGTCGCCTTCACGCCCAACGCCCTGCAGGCCATGAGAGTC 180
Transcript    ATGCAAGAAATCGGCGCCGGGGTCGCCTTCACGCCCAACGCCCTGCAGGCCATGAGAGTC 180
GO851910      ------------------------------------------------------------ 0
GO814712      ------------------------------------------------------------ 0
GO814910      ------------------------------------------------------------ 0
GO805203      ------------------------------------------------------------ 0
GO805370      ------------------------------------------------------------ 0

Gene          TGCCATCCGGCCATATACGAGGCCTACGAGAGGGTGCGCACGCGCAACCTGTGGCCCTGC 240
Transcript    TGCCATCCGGCCATATACGAGGCCTACGAGAGGGTGCGCACGCGCAACCTGTGGCCCTGC 240
GO851910      ------------------------------------------------------------ 0
GO814712      ------------------------------------------------------------ 0
GO814910      ------------------------------------------------------------ 0
GO805203      ------------------------------------------------------------ 0
GO805370      ------------------------------------------------------------ 0

Gene          AAGCAAAAGGTATGGTTCGACTTCTACGACACGCGGAACGAGGATGGTGATGCGGCTGCC 300
Transcript    AAGCAAAAGGTATGGTTCGACTTCTACGACACGCGGAACGAGGATGGTGATGCGGCTGCC 300
GO851910      ------------------------------------------------------------ 0
GO814712      ------------------------------------------------------------ 0
GO814910      ------------------------------------------------------------ 0
GO805203      ------------------------------------------------------------ 0
GO805370      ------------------------------------------------------------ 0

Gene          AAGCCTGCCTTTACCATCTCAAACGACGTCGGGCAGAACGGGGTCCTTCGGGCGCATTTC 360
Transcript    AAGCCTGCCTTTACCATCTCAAACGACGTCGGGCAGAACGGGGTCCTTCGGGCGCATTTC 360
GO851910      ------------------------------------------------------------ 0
GO814712      ------------------------------------------------------------ 0
GO814910      ------------------------------------------------------------ 0
GO805203      ------------------------------------------------------------ 0
GO805370      ------------------------------------------------------------ 0
```

```
Gene         CTGGATGAGCTCATCAAGCTGGTGCCGAGGGAGGCTGCGCACGTCGGGAAGAGGCTGGCG  420
Transcript   CTGGATGAGCTCATCAAGCTGGTGCCGAGGGAGGCTGCGCACGTCGGGAAGAGGCTGGCG  420
GO851910     ------------------------------------------------------------  0
GO814712     ------------------------------------------------------------  0
GO814910     ------------------------------------------------------------  0
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         AGCTACGAGGAGGGGGGCCCCGACGGGAGACTGCGACTCAGGTTCGCCGATGGCTCGGAG  480
Transcript   AGCTACGAGGAGGGGGGCCCCGACGGGAGACTGCGACTCAGGTTCGCCGATGGCTCGGAG  480
GO851910     ------------------------------------------------------------  0
GO814712     ------------------------------------------------------------  0
GO814910     ------------------------------------------------------------  0
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         GACGAGGCGGATGTCATTCTGGCCTGTGACGGGATCAAGTCCAGGGTCCGCCAGCTGCTG  540
Transcript   GACGAGGCGGATGTCATTCTGGCCTGTGACGGGATCAAGTCCAGGGTCCGCCAGCTGCTG  540
GO851910     ------------------------------------------------------------  0
GO814712     ------------------------------------------------------------  0
GO814910     ------------------------------------------------------------  0
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         TTCGGCGCCCATCATCCTTGTGCTCTCCCGTCGTATACCCACAGGTACGCGTATCGAGCT  600
Transcript   TTCGGCGCCCATCATCCTTGTGCTCTCCCGTCGTATACCCACAGGTACGCGTATCGAGCT  600
GO851910     ------------------------------------------------------------  0
GO814712     ------------------------------------------------------------  0
GO814910     ------------------------------------------------------------  0
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         CTGGTGCCGATGGAAGAGGCCGTCGACGCCATCGGGAAAGAGAAGGCCCAGAATGCAGCC  660
Transcript   CTGGTGCCGATGGAAGAGGCCGTCGACGCCATCGGGAAAGAGAAGGCCCAGAATGCAGCC  660
GO851910     ------------------------------------------------------------  0
GO814712     ------------------------------------------------------------  0
GO814910     ------------------------------------------------------------  0
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         ATGCACGTATGTAAGGCTTCAAAACAAGCTTTTTGTCCCACCCCCACGAAACACGCACAC  720
Transcript   ATGCAC------------------------------------------------------  665
GO851910     ------------------------------------------------------------  0
GO814712     ------------------------------------------------------------  0
GO814910     ------------------------------------------------------------  0
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         CAAACCCGTTTTCAAATTCCCCCGTCACCGTCACCGTCACCGTCACCTTTTGTCC----C  776
Transcript   ------------------------------------------------------------  665
GO851910     ----------------------------CGCCCGGGGGACCGTCACCTTTTGTGCCAGTC  32
GO814712     ----------------------------------------ACCGTCACCTTTTGTCCCAGTC  22
GO814910     ---------------------------------GGGGGACCGTCACCTTTTGTCCCAGTC  27
GO805203     ------------------------------------------------------------  0
GO805370     ------------------------------------------------------------  0

Gene         AGTCAGTCAGTCAGTCAGTCAAGTTACTGTTGATGCGGAAAAGCCACTGACATTGTCTGC  836
Transcript   ------------------------------------------------------------  665
```

```
GO851910     AGTCAGTCAGTCAGTCAGTCAAGTTACCGTTGATGCGGAAGAGCCACTGACATTGTTTGT  92
GO814712     AGTCAGTCAGTCAGTCAGTCAAGTTACCGTTGATGCGGAAAAGCCACTGACATTGTCTGC  82
GO814910     AGTCAGTCAGTCAGTCAAGTTACCGTTGATGCGGAAAAGCCACTGACATTGTCTGC  87
GO805203     AGTCAGTCAGTCAGTCAGTCAAGTTACCGTTGATGCGGAAAAGCCACTGACATTGTCTGC  60
GO805370     ------------------------------------------------------TTGTCTGT  8

Gene         TACGACAAAAAAAAAAAAAACGTCT-ACAGATGGGCAAAGGCGGCCACGTTTTGACCTTCC  896
Transcript   -----------------------------ATGGGCAAAGGCGGCCACGTTTTGACCTTCC  697
GO851910     TACGACAAAAAAAAAAA---CGTCT-ACAGATGGGCAAAGGCGGCCACGTTTTGACCTTCC  149
GO814712     TACGACAAAAAAAAAAA---CGTCT-ACAGATGGGCAAAGGCGGCCACGTTTTGACCTTCC  139
GO814910     TACGACAAAAAAAAAAA---CGTCT-ACAGATGGGCAAAGGCGGCCACGTTTTGACCTTCC  144
GO805203     TACGACAAAAAAAAAAA---CGTCT-ACAGATGGGCAAAGGCGGCCACGTTTTGACCTTCC  117
GO805370     TACGACAAAAAAAAAAAA--CGTCTTACAGATGG-CAAAGGCGGCCACGTTTTGACCTTCC  66
                                          **** **************************

Gene         CTGTCAACCACGGGCAGACAGTCAAT-GTCGTTGCGTTCCACGCAACCTCGGAGGAATGGC  956
Transcript   CTGTCAACCACGGGCAGACAGTCAAT-GTCGTTGCGTTCCACGCAACCTCGGAGGAATGGC  757
GO851910     CTGTCAACCACGGGCAGACAATCAAT-GTCGTTGCGTTCCACGCAACCTCGGAAGAATGGC  209
GO814712     CTGTCAACCACGGGCAGACAGTCAAT-GTCGTTGCGTTCCACGCAACCTCGGAGGAATGGC  199
GO814910     CTGTCAACCACGGGCAGACAGTCAAT-GTCGTTGCGTTCCACGCAACCTCGGAGGAATGGC  204
GO805203     CTGTCAACCACGGGCAGACAGTCAAT-GTCGTTGCGTTCCACGCAACCTCGGAGGAATGGC  177
GO805370     -TGTCAACCACGGGCAGACAGTCAATTGTCGTTGCGTTCCACGCAACCTCGGAGGAATGGC  126
              ************************* ***********************************

Gene         CGGACAGCTCAAAGCTCACCGCGCAATCTACCCGTGAGGCGGCGCTGCGGGACTTTTCCA  1016
Transcript   CGGACAGCTCAAAGCTCACCGCGCAATCTACCCGTGAGGCGGCGCTGCGGGACTTTTCCA  817
GO851910     CGGACAACTCAAAGCTAACCGCGCAATCTACCCGTGAGGGGGCGTTGCGGGACTTTTCCA  269
GO814712     CGGACAGCTCAAAGCTAACCGCGCAATCTACCCGTGAGGCGGCGCTGCGGGACTTTTCCA  259
GO814910     CGGACAGCTCAAAGCTAACCGCGCAATCTACCCGTGAGGCGGCGCTGCGGGACTTTTCCA  264
GO805203     CGGACAGCTCAAAGCTAACCGCGCAATCTACCCGTGAGGCGGCGCTGCGGGACTTTTCCA  237
GO805370     CGGACAGCTCAAAGCTAACCGCGCAATCTACCCGTGAGACGGCGCTGCGGGACTTTTCCA  186
             ****** ********* *********************  **** ***************

Gene         ACTTCGGACCCGGAATCACCAAGCTGCTGAAGTTGACGAGCCCGAAACTCGACATTGTGA  1076
Transcript   ACTTCGGACCCGGAATCACCAAGCTGCTGAAGTTGACGAGCCCGAAACTCGACATT----  871
GO851910     ACTTTGGACCCGGAATCACC-------------------------------------  289
GO814712     ACTTCGGACCCGGAATCACC-------------------------------------  279
GO814910     ACTTTGGACCCGGAATCACC-------------------------------------  284
GO805203     ACTTCGGACCCGGAATCACCAAGCTGCTGAAGTTGACGAGCCCGAAACTCGACATT----  292
GO805370     ACTTCGGACCCGGAATCACCAAGCTGCTGAAGTTGACGAGCCCGAAACTCGACATT----  241
             **** ***************

Gene         GTTATTGCCCCTTTTCCCCTTGTTTGAGGCGCAGCGCAGCGCAGCAGAAAAAAAAAAACA  1136
Transcript   ------------------------------------------------------------  871
GO851910     ------------------------------------------------------------  289
GO814712     ------------------------------------------------------------  279
GO814910     ------------------------------------------------------------  284
GO805203     ------------------------------------------------------------  292
GO805370     ------------------------------------------------------------  241

Gene         AAAACAAACAAAAAACCAAAAAAAACATACATTGGGGAGAAGAACAAGAACATGAATTGA  1196
Transcript   ------------------------------------------------------------  871
GO851910     ------------------------------------------------------------  289
GO814712     ------------------------------------------------------------  279
GO814910     ------------------------------------------------------------  284
GO805203     ------------------------------------------------------------  292
GO805370     ------------------------------------------------------------  241
```

```
Gene         TGATGATGACAAAATTTTCAAAGTGGGCAATCTTCGATTTGGGCGAGAATCCCCCGCCCA 1256
Transcript   ----------------------TGGGCAATCTTCGATTTGGGCGAGAATCCCCCGCCCA 910
GO851910     ------------------------------------------------------------A 290
GO814712     ------------------------------------------------------------A 280
GO814910     ------------------------------------------------------------A 285
GO805203     ----------------------TGGGCAATCTTCGATTTGGGCGAGAATCCCCCGCCCA 330
GO805370     ----------------------TGGGCAATCTTCGATTTGGGCGAGAATCCCCCGCCCA 279
                                                                        *

Gene         CGTTTGCCATGGATCGCGTCTGTCTGGTCGGCGACGCCGCCCACGCGACGTCCCCGCATC 1316
Transcript   CGTTTGCCATGGATCGCGTCTGTCTGGTCGGCGACGCCGCCCACGCGACGTCCCCGCATC 970
GO851910     CGTTTGCCATGGATCGCGTCTGTCTGGTCGGGGACGCCGCCCCCGCGACGTCCCCCCATC 350
GO814712     CGTTTGCCATGGATCGCGTCTGTCTGGTCGGCGACGCCGCCCACGCGACGTCCCCGCATC 340
GO814910     CGTTTGCCATGGATCGCGTCTGTTTGGTCGGGGACGCCGCCCCCGCGACGTCCCCGCATC 345
GO805203     CGTTTGCCATGGATCGCGTCTGTCTGGTCGGCGACGCCGCCCACGCGACGTCCCCGCATC 390
GO805370     CGTTTGCCATGGATCGCGTCTGTCTGGTCGGCGACGCCGCCCACGCGACGTCCCCGCATC 339
             ********************** ******* ********** ************ ****

Gene         ATGGGGCCGGGGCGGGATTCTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 1376
Transcript   ATGGGGCCGGGGCGGGATTCTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 1030
GO851910     ATGGGGCCGGGGCGGGATTTTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 410
GO814712     ATGGGGCCGGGGCGGGATTCTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 400
GO814910     ATGGGGCCGGGGCGGGATTTTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 405
GO805203     ATGGGGCCGGGGCGGGATTCTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 450
GO805370     ATGGGGCCGGGGCGGGATTTTGCATCGAGGACGCCGCGGTGCTGGCGCACTTGCTCGCCA 399
             ******************** ****************************************

Gene         GCGAGGAAATCTCGGATCATCGCGGCCACCGGGTGGCTCTTGCGGTGTACGACGCCGTGA 1436
Transcript   GCGAGGAAATCTCGGATCATCGCGGCCACCGGGTGGCTCTTGCGGTGTACGACGCCGTGA 1090
GO851910     GCGAGGAAATCTCGGATCATCGCGGCCCCCGGGTGGCTCTTGCCGTGTACGACGCCGTGA 470
GO814712     GCGAGGAAATCTCGGATCATCGCGGCCACCGGGTGGCTCTTGCGGTGTACGACGCCGTGA 460
GO814910     GCGAGGAAATCTCGGATCATCGCGGCCCCCGGGTGGCTCTTGCGGTGTACGACGCCGTGA 465
GO805203     GCGAGGAAATCTCGGATCATCGCGGCCACCGGGTGGCTCTTGCGGTGTACGACGCCGTGA 510
GO805370     GCGAGGAAATCTCGGATCATCGCGGCCACCGGGTGGCTCTTGCGGTGTACGACGCCGTGA 459
             *************************** *************** ****************

Gene         GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACACGTACGAGT 1496
Transcript   GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACACGTACGAGT 1150
GO851910     GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACCCGTACGAGT 530
GO814712     GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACACGTACGAGT 520
GO814910     GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACACGTACGAGT 525
GO805203     GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACACGTACGAGT 570
GO805370     GACGAGAGCGTGGCTCGTGGTTGGTGCAGAGCAGTAGACGCATCGGCGACACGTACGAGT 519
             ************************************************* *********

Gene         GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 1556
Transcript   GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 1210
GO851910     GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 590
GO814712     GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 580
GO814910     GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 585
GO805203     GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 630
GO805370     GGATGGCCGAGGGGATAGAGGACGACCTGGCGAAGGCTGAGGAGGAGATCAAGTATAGGA 579
             ************************************************************

Gene         ACGGGGTGATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT 1616
Transcript   ACGGGGTGATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT 1270
GO851910     ACGGGGTAATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT 650
GO814712     ACGGGGTAATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT 640
GO814910     ACGGGGTAATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT 645
```

```
GO805203      ACGGGGTAATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT  690
GO805370      ACGGGGTAATTGCCGACGTGGATGTCGAGGCCATGTGTCATCAGGCGCGAGAGGAGTTTT  639
              ******* **************************************************

Gene          CGAAAAGGTGGCGTGCCGTTGGTTAG---------------------------------  1642
Transcript    CGAAAAGGTGGCGTGCCGTTGGTTAG---------------------------------  1296
GO851910      CCAAAAGGGGGCGGGCCGTTGGTTAGTTAATTGAATGATTTGATTGATTCTCTCCCGGAA  710
GO814712      CCAAAAGGTGGCGTGCCGTTGGTTAGTTAATTGATTGATTTGATTGATTCTCTCCCGGAA  700
GO814910      CCAAAAGGGGGCGTGCCGTTGGTTAGTTAATTGATTGATTTGATTGATTCTCTCCCGGAA  705
GO805203      CCAAAAGGTGGCGTGCCGTTGGTTAGTTAATTGATTGATTTGATT-----------GGAA  750
GO805370      CCAAAAGGTGGCGTGCCGTTGGTTAGTTAATTGATTGATTTGATT-----------GGAA  699
              * ****** **** ************

Gene          -------------------------  1642
Transcript    -------------------------  1296
GO851910      AAAAAAAAAAAAAAAAAAAAAAAAA  735
GO814712      AAAAAAAAAAAAAAAAAAAAAAAAA  725
GO814910      AAAAAAAAAAAAAAAAAAAAAAAAA  730
GO805203      ----------AAAAAAAAAAAAAAA  754
GO805370      ----------AAAAAAAAAAAAAAA  703
```

Legend

Exon 1 | Exon 2 | Exon 3

Expression and purification of *E. festucae* Rose City salicylate hydroxylase protein

To study the expression of *E. festucae* RC salicylate hydroxylase, the coding sequence of the putative functional transcript was cloned into the expression vector pET21b. The recombinant plasmid was grown in *E. coli* BL21-CodonPlus (DE3)-RIPL expression strain. Expression of the recombinant salicylate hydroxylase gene was confirmed by performing colorimetric enzyme assays using salicylic acid (Cane and Williams, 1982). The heterologously expressed protein was observed to degrade the salicylic acid in the nutrient media. The product of salicylate hydroxylase activity is catechol, which undergoes oxidation producing a brown color (Figure 4.5).

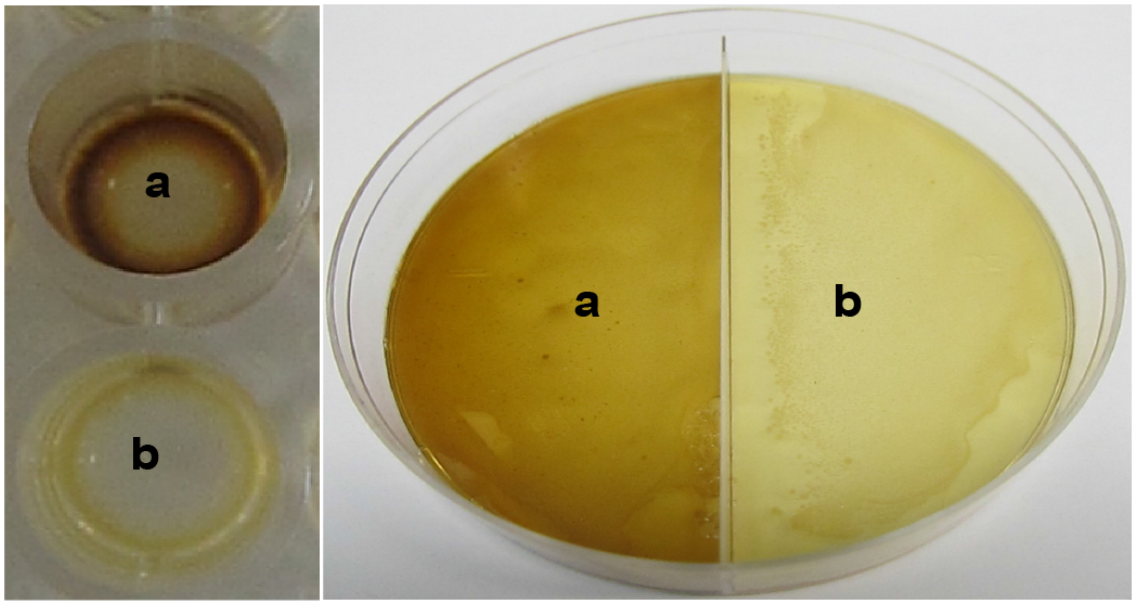


Figure 4.5. Colorimetric assay to confirm the heterologous expression of *E. festucae* Rose City salicylate hydroxylase. (**a**) *E. coli* cells expressing *E. festucae* Rose City salicylate hydroxylase on nutrient liquid media in a microtiter plate (left), and solid media (right) were able to utilize salicylate as a substrate as evidenced by the formation and accumulation of catechol, a byproduct of the reaction that turns dark brown upon auto-oxidation. (**b**) Bacterial cells expressing the pET21b vector only used as negative control retain the original pale color of the LB medium.

Once recombinant gene expression was confirmed, attempts to purify the salicylate hydroxylase protein expressed in *E. coli* were made. Initial trials to obtain appreciable amounts of protein in the soluble phase of the bacterium were made difficult by the repeated localization of the overexpressed protein in inclusion bodies. Despite several modifications to aid the solubility of the expressed protein, such as lowering induction temperatures to favor proper folding of the native protein (Strandberg and Enfors, 1991; Lilie et al., 1998; Suzuki et al., 2000; Zhao et al., 2005; Esposito and Chatterjee, 2006), and reducing IPTG amounts to slow the rate of the recombinant protein synthesis to reduce accumulation in the bacterial cytoplasm (Kopetzki et al., 1989; Galloway et al., 2003), the induced protein still was expressed as insoluble aggregates

(Figure 4.6 Lanes 1-3). Similar hurdles were reported by authors of previous studies who sought to overexpress Sal, NahU, and NahG, the salicylate hydroxylases found in *Pseudomonas* spp. (Suzuki et al., 2000; Zhao et al., 2005). Although in these cases, lowering the induction temperature from 37°C to 20°C easily solubilized the bacterial protein.

Sufficient amounts of soluble homologous salicylate hydroxylase were finally obtained by expressing the protein in ArcticExpress RP (DE) bacterial cells (Figure 4.6 Lanes 4-6). This strain of *E. coli* was engineered to co-expresses Cpn10 and Cpn60, the cold-adapted chaperonins from the psychrophilic bacterium, *Oleispira antarctica* (Ferrer et al., 2003). Since the *O. antarctica* chaperonins have high similarity to the native *E. coli* GroEL and GroES chaperonins, the ArcticExpress RP (DE) bacterial cells were reported to improve protein processing and folding at lower temperatures, therefore helping to increase the amount of soluble heterologous protein expression (Hartinger et al., 2010; Lin et al., 2012; Pacheco et al., 2012). Although soluble expression of the *E. festucae* RC salicylate hydroxylase was still low, the yield was adequate for further downstream applications.

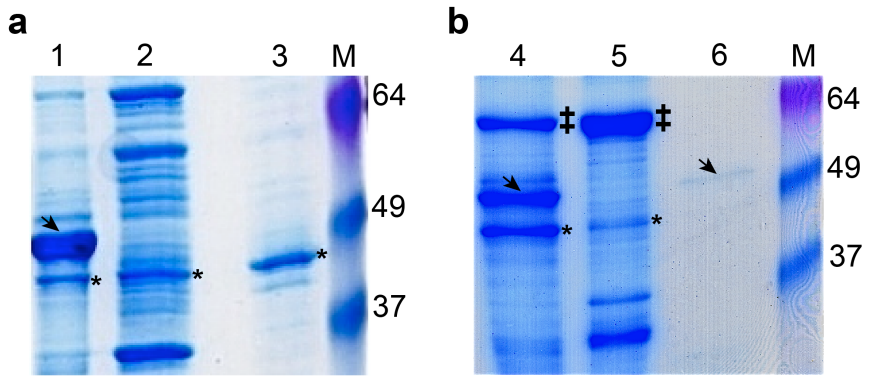


Figure 4.6. Expression of *E. festucae* Rose City salicylate hydroxylase in two *E. coli* expression strains. SDS-PAGE analysis of *E. festucae* Rose City salicylate hydroxylase expressed in *E. coli* (**a**) BL21 CodonPlus RIPL (DE) cells showing Lane 1: insoluble pellet, Lane 2: 25µg total soluble crude, and Lane 3: nickel-ion magnetic beads bound eluted proteins using MagneHis Protein Purification System consisting mainly of the contaminating bacterial protein SlyD, (**b**) ArcticExpress RP (DE) cells showing Lane 4: insoluble pellet, Lane 5: 25µg total soluble crude, and Lane 6: 0.5 µg purified protein using TALON Metal Affinity Resin.
Arrow indicates presence of the induced protein at the expected size of 48.3 kD. Asterisk is used to show SlyD, the 44 kD nickel-binding *E. coli* dimer protein (see below for further discussion). The 60 kD bacterial chaperonin protein, Cpn60 expressed in *E. coli* ArcticExpress cells is indicated by ‡. Standard protein markers are indicated by M.

Detection of a persistent *Escherichia coli* metal-binding contaminating protein

Initial attempts to purify the recombinant *E. festucae* RC salicylate hydroxylase were done using nickel-nitrilotriacetic acid (Ni^{2+}- NTA). MagneHis Ni^{2+}- NTA magnetic resin bead purification of *E. coli* cells expressing the salicylate hydroxylase protein resulted in a prominent band between 40 – 50 kDalton upon SDS-PAGE (Figure 4.6a, Lane 3). Since the band's placement was slightly lower than the expected 48.3 kDalton of the polyHis-tagged *E. festucae* RC salicylate hydroxylase protein, *E. coli* cells expressing the pET21b vector control only was induced with

IPTG alongside the cells expressing the salicylate hydroxylase protein, and visualized simultaneously. Surprisingly, the same band was also observed in the *E. coli* cells expressing vector control only (data not shown), indicating the presence of a metal-binding bacterial protein that was bound to, and eluted from the Ni^{2+}- NTA magnetic beads.

Literature search revealed that the nuisance protein most likely was the *E. coli* peptidyl proline *cis–trans* isomerase, Sensitivity to lysis, SlyD (Maratea et al., 1985; Roof et al., 1994). Mitterauer et al. (1999) reported that SlyD migrated as a 44 kDalton dimer. This was similar to the size of the band that was observed in both the purified *E. coli* cells expressing the salicylate hydroxylase protein, and vector control only. SlyD's histidine rich C-terminal metal binding domain (Scholz et al., 2006), and the presence of the sequence ^{N}His-Asp-His-His-His-Asp-His^{C} (Wülfing et al., 1994) was postulated to allow chelating to Ni^{2+}- NTA (Mukherjee et al., 2003). Consequently SlyD was recurrently co-eluted during Ni^{2+}- NTA purification of heterologous polyHis-tagged proteins, even more so when recombinant protein expression is low (Robichon et al., 2011; Finzi et al., 2003).

Purified soluble *E. festucae* RC salicylate hydroxylase was obtained only when TALON Metal Affinity Resin, a cobalt ion-charged immobilized metal ion affinity chromatography (IMAC), was applied to the soluble protein fraction. MagneHis nickel Ni^{2+}- NTA resin purification of the recombinant protein repeatedly yielded a bacterial contaminant, SlyD that the TALON Metal Affinity Resin did not co-purify (McMurry and Macnab, 2004). Cobalt was reported to have higher spatial binding specificity to polyHis-tagged heterologous proteins relative to nickel in addition to low metal ion leakage (Chaga et al., 1999; Chaga et al., 1999) thus aiding in the recovery of purified recombinant protein (Phan et al., 2003; Antaloae et al., 2013). The purification of the heterologous *E. festucae* RC was achieved by switching to an *E. coli* strain that facilitates soluble protein expression, and using cobalt-charged resin for affinity-tag purification.

Catalytic activity of *E. festucae* Rose City salicylate hydroxylase

The purified *E. festucae* RC salicylate hydroxylase was able to metabolize salicylic acid, and 2-5-dihydroxybenzoic acid (gentisate). The latter was the more efficient substrate for the fungal endophyte enzyme (Table 4.2). The activity of *E. festucae* RC salicylate hydroxylase activity with 2-5-dihydroxybenzoic acid as substrate was tested because Szu and Dagley (1984) reported that salicylate hydroxylase from the fungus *Trichosporon cutaneum* exhibited three times higher activity rate with the compound relative to salicylic acid. The activity rate was also higher for the fungal endophyte enzyme, although not three-fold. In *P. putida*, Balashova et al. (2001) showed that the activity of the enzyme with 2-5-dihydroxybenzoic acid was 114% relative to salicylate.

Table 4.2. Mean activity of triplicate assays obtained using purified heterologously expressed *E. festucae* RC salicylate hydroxylase. Specific activities are measured in nanomoles NADH minute^{-1} milligram protein^{-1}.

Substrate	Mean Activity	Standard deviation	Relative activities (%)
Salicylate	713.34	75.06	100
Gentisate	1,038.34	68.25	146

The *E. festucae* RC salicylate hydroxylase activity with salicylic acid as the substrate was observed to be higher relative to the reported *U. maydis* GST-Shy1 activity (300 nmol NADH min^{-1} mg protein^{-1}) and *Pseudomonas* spp. GST-NahG (550 nmol NADH min^{-1} mg protein^{-1}) for the same compound (Rabe et al., 2013). However, Zhao et al. (2005) reported a much higher activity for purified *Pseudomonas* spp. NahG with salicylic acid as the substrate (3,170 nmol NADH min^{-1} mg protein^{-1}) and markedly more for NahU (23,440 nmol NADH min^{-1} mg protein^{-1}). The NahG activity by Rabe et al. (2013) may have been lower due to the uncleaved 26 kDalton GST tag fused to the protein. The presence of relatively large affinity purification and

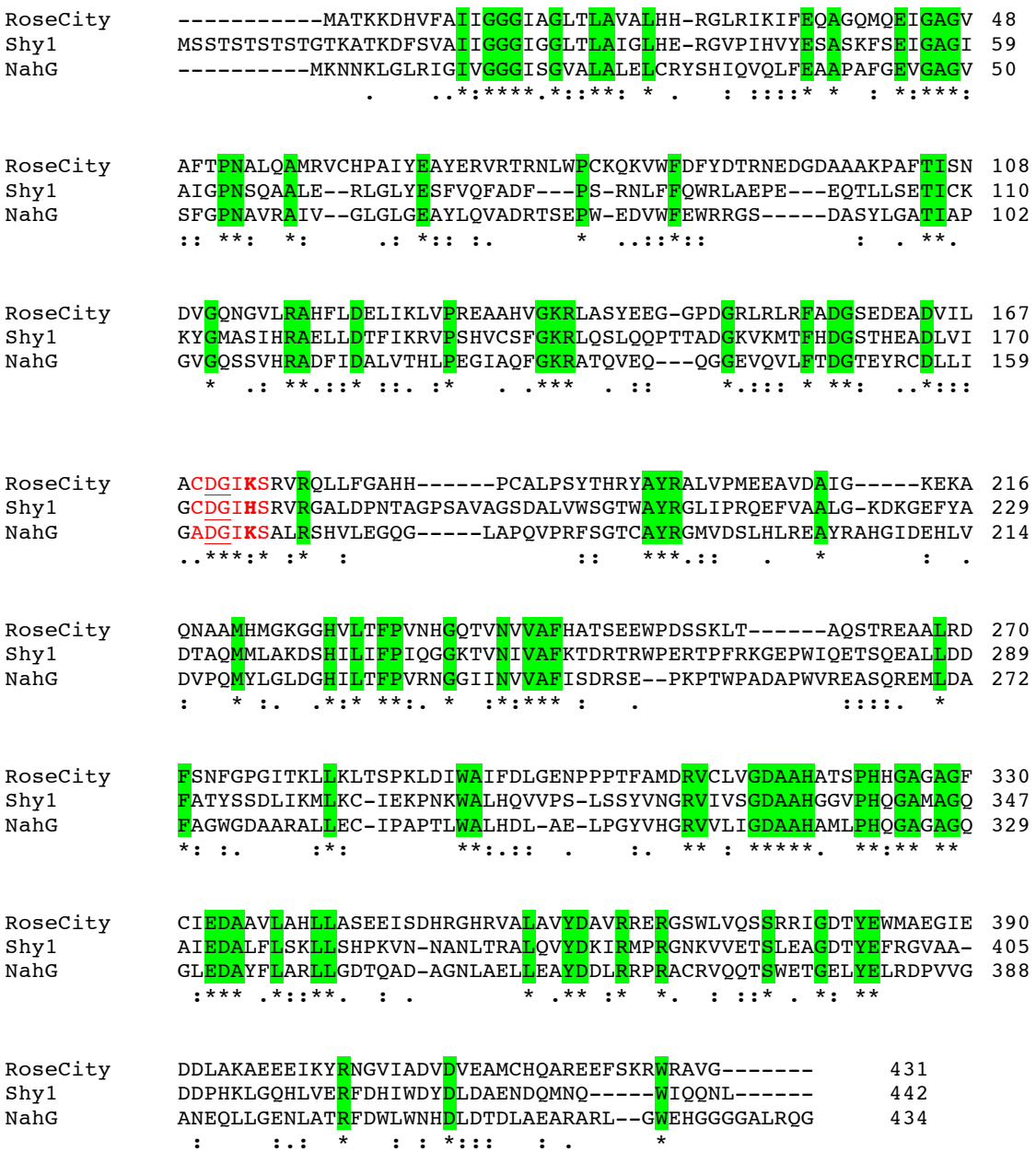
solubility-enhancing fusion tags like GST may impact the functional and structural properties of the recombinant protein (Purbey et al., 2006). Hence it is highly desirable that GST tags are cleaved off the protein prior to downstream biochemical assays (Esposito and Chatterjee, 2006).

Discussion

The successful cloning of the gene encoding salicylate hydroxylase from *E. festucae* Rose City leading to the expression, purification and functional characterization of the salicylate hydroxylase enzyme is described in this chapter. The gene was experimentally annotated to be 1,642 bp, and transcribed a functional 1,296 bp transcript that comprised three exons which were interrupted by two introns. The encoded salicylate hydroxylase protein consisted of 431 amino acids that migrated as a 48.3 kDalton protein during SDS-PAGE.

Amino acid alignment of the fungal endophyte salicylate hydroxylase sequence with the classic *P. putida* NahG, and *U. maydis* Shy1, the only cloned and characterized fungal salicylate hydroxylase to date, revealed high similarity among all three proteins, and overall conservation of the NAD(P)H and FAD binding domains (Figure 4.7). Interestingly, the lysine residue that has been reported to be critical for NAD(P)H affinity (Suzuki et al., 2000) is present in the *E. festucae* RC salicylate hydroxylase as in NahG. In Shy1, a histidine residue is substituted in the position (Rabe et al., 2013) similar to the NAD(P)H binding site of *P. fluorescens* p-hydroxybenzoate hydroxylase, PHBH (Eppink et al., 1997; Eppink et al., 1998), a model enzyme for the study of oxygenation of an aromatic substrate (Suemori, 2013).

Figure 4.7. Amino acid alignment of salicylate hydroxylase sequences from *E. festucae* Rose City (this study), *U. maydis* Shy1 (XP_761377), and *P. putida* NahG (P23262). Identical residues are shown in green. The conserved site that is involved in NAD(P)H binding including the crucial lysine, K/ histidine, H (Eppink et al., 1997; Suzuki et al., 2000; Rabe et al., 2013) residues are highlighted in bold red, while the conserved motif for FAD binding, DG is underlined (Eppink et al., 1997).



Transcript variation of the gene seemed to be prevalent as observed by the presence of at least two different forms of spliced transcripts that were determined to not form a functional protein. Expressed sequence tags available from NCBI for the gene from the *E. festucae* isolate of meadow fescue revealed additional similar transcript variants. The mechanism of splicing as a whole, and alternative splicing especially, is consistently invoked to account for the modulation, and tissue specificity of gene expression in organisms (Matlin et al., 2005; Keren et al., 2010). Alternative splicing also plays an important role in the diversification of proteomes, and is involved in gene expression patterns that are unique to a given species (Kornblihtt et al., 2013).

In fungi, it is uncommon for alternative splicing of transcripts to give rise to functional proteins, and the occurrence of intron retention may be important in regulating gene expression (Marshall et al., 2013). McGuire et al. (2008) reported that the type of alternative splicing described here, where segments of introns were retained, was seen predominantly in fungi and protists relative to other eukaryotes in their comparison of alternative splicing and splice recognition patterns across kingdoms. The authors stated that retained introns, a natural and frequent phenomenon, tend to result from inconsistent splice site recognitions of intron-exon boundaries due to failure of the internal intron definition mechanism. Most transcript variants with retained introns are unlikely to translate into a protein due to an early termination site or eventual reading frameshift (McGuire et al., 2008), consistent with the observation in *E. festucae* salicylate hydroxylase.

As described earlier, the persistent and recalcitrant IMAC-contaminating *E. coli* protein confounded efforts to obtain appreciable amounts of pure heterologous *E. festucae* RC salicylate hydroxylase protein. The discovery of SlyD occurred during an investigation of bacteriophage φX174's E-protein activity that induced cell lysis in *E. coli* (Maratea et al., 1985). SlyD is a host factor needed for phage lysis (Maratea et al., 1985) and also functions as a nucleotide binding protein (Mitterauer et al., 1999). Almost a decade later, the protein was independently found

because of its persistent contamination of IMACs during purification of heterologous proteins, and given the name Wondrous Histidine-rich Protein, WHP (Wülfing et al., 1994). Unexpectedly, the protein is able to retain its strong affinity to metals, especially nickel and zinc, even under denaturing conditions (Wülfing et al., 1994; Mukherjee et al., 2003).

It is most likely that SlyD contamination of MagneHis Ni^{2+}- NTA during purification attempts of the recombinant *E. festucae* RC salicylate hydroxylase stemmed primarily from low levels of soluble protein. The induced salicylate hydroxylase protein was found, almost entirely, aggregated in the insoluble fraction and only low amounts were detected in the soluble phase as described previously (Figure 4.3). Parsy et al. (2007) reported that SlyD out-competed their heterologous polyHis-tagged protein for binding to Ni^{2+}- NTA matrix.

It was also reported that in cases of low recombinant protein expression, metal binding bacterial proteins like SlyD are purified in equal or greater amounts than the original polyHis-tagged protein (Andersen et al., 2013). In cases where there was little size difference between the bacterial contaminant and the protein of interest, in addition to high expression of the contaminant as described in this study, it was reported that the contaminant metal-binding bacterial protein might be purified in place of the recombinant protein (Bolanos-Garcia and Davies, 2006). Lichty et al. (2005) described the co-purification of 43 kDalton *E. coli* contaminant protein during the purification of a 44.6 kDalton weakly-expressed recombinant protein that was also prone to form insoluble aggregates. Taken together, it is likely that the polyHis-tagged recombinant *E. festucae* RC salicylate hydroxylase protein was outcompeted by SlyD's strong affinity for Ni^{2+}- NTA owing to the low soluble heterologous protein yield.

A troublesome lack of soluble protein expression that is reported here might have been also encountered during the characterization of Shy1, an active salicylate hydroxylase found in the

biotrophic fungus, *Ustilago maydis* (Rabe et al., 2013). Although no specific mention of such problems was made, they reported that glutathione-S-transferase (GST) fusion of Shy1 was generated to optimize protein expression and solubility, whereas His-tagged fusions were used for the inactive salicylate hydroxylase homologs of *U. maydis*. Furthermore there were substantially lower amounts of the purified GST-Shy1 protein shown on SDS-PAGE relative to the His-tagged inactive protein using identical treatments. However as purified protein yields were not reported in the study, it is unclear how much GST-Shy1 was recovered in soluble form.

The degradation of salicylate by bacterial cells expressing *E. festucae* RC salicylate hydroxylase on nutrient media, with subsequent auto-oxidation of accumulated catechol that was characterized by an unmistakable dark brown saturation of media, confirmed that the gene responsible for salicylate hydroxylase was cloned and expressed heterologously. The catalytic assays of the purified protein indicated that the enzyme is active with salicylate and gentisate as substrates, similar to the well-studied *Pseudomonas* spp. salicylate hydroxylases and the enzyme purified from *T. cutaneum*. The standard NADH assay was used to measure the activity of the enzyme since the oxidation of NADH to NAD^+ could be monitored by the corresponding decrease in absorbance at $\lambda = 340$ nm. This decrease is stoichiometric with the degradation of salicylate and the concomitant formation of the by-product catechol (Figure 4.1). The activity of the enzyme with gentisate was higher relative to salicylate, a trend consistent with the salicylate hydroxylases from *T. cutaneum* and *P. putida*.

Rabe et al. (2013) presented evidence for *U. maydis* to use salicylic acid as a sole carbon source for growth on minimal media. Surprisingly, their work showed that Shy1 was not required for the pathogen's virulence during maize seedling infection although transcript up-regulation exceeded two orders of magnitude during various infection stages. The authors concluded that the results might be due to the presence of additional genes and pathways, which are not detectable in axenic

culture studies, that degrade salicylic acid as a plant defense hormone. An insight into possible roles of the *E. festucae* RC salicylate hydroxylase enzyme could be obtained depending on whether the fungal endophyte could utilize salicylic acid as a carbon source, as seen in *U. maydis* (Rabe et al. 2013) and *Fusarium* sp. (Dodge and Wackett, 2005).

The work described in this chapter encompassed the cloning, expression, purification, and catalytic activity determination of a fungal salicylate hydroxylase. Until now, there is no documented work on salicylate hydroxylases of fungal symbionts of plants. Hence the study discussed here is a first in the attempt to understand the function of this enzyme in the complex plant-fungal beneficial endosymbiont interplay. Whether *E. festucae* RC salicylate hydroxylase is a factor in the evasion of host defenses is not yet known.

References

1. Ambrose KV, Belanger FC (2012) SOLiD-SAGE of endophyte-infected red fescue reveals numerous effects on host transcriptome and an abundance of highly expressed fungal secreted proteins. PLoS ONE, 7:e53214.

2. Andersen KR, Leksa NC, Schwartz TU (2013) Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His‑tag purification. Proteins: Structure, Function, and Bioinformatics, 81:1857-1861.

3. Antaloae AV, Montigny C, le Maire M, Watson KA, Sørensen TLM (2013) Optimisation of recombinant production of active human cardiac SERCA2a ATPase. PLoS ONE, 8: e71842.

4. Balashova NV, Stolz A, Knackmuss HJ, Kosheleva IA, Naumov AV, Boronin AM (2001) Purification and characterization of a salicylate hydroxylase involved in 1-hydroxy-2-naphthoic acid hydroxylation from the naphthalene and phenanthrene-degrading bacterial strain *Pseudomonas putida* BS202-P1. Biodegradation, 12:179-188.

5. Bolanos-Garcia VM, Davies OR (2006) Structural analysis and classification of native proteins from *E. coli* commonly co-purified by immobilised metal affinity chromatography. Biochimica et Biophysica Acta, 9:1304-1313.

6. Bosch R, Moore ER, García-Valdés E, Pieper DH (1999) NahW, a novel, inducible salicylate hydroxylase involved in mineralization of naphthalene by *Pseudomonas stutzeri* AN10. Journal of Bacteriology, 181:2315-2322.

7. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72:248-254.

8. Camara B, Bielecki P, Kaminski F, dos Santos VM, Plumeier I, Nikodem P, Pieper DH (2007) A gene cluster involved in degradation of substituted salicylates via ortho cleavage in *Pseudomonas* sp. strain MT1 encodes enzymes specifically adapted for transformation of 4-methylcatechol and 3-methylmuconate. Journal of Bacteriology, 189:1664-74.

9. Cane PA, Williams PA (1982) The plasmid-coded metabolism of naphthalene and 2-methyl-naphthalene in *Pseudomonas* strains: phenotypic changes correlated with structural modification of the plasmid pWW60-1. Journal of General Microbiology, 128:2281-2290.

10. Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M, Paley S, Rhee SY, Shearer AG, Tissier C, Zhang P, Karp PD (2008) The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. Nucleic Acids Research, 36:D623-D631.

11. Chaga G*,* Bochkariov DE*,* Jokhadze GG*,* Hopp J*,* Nelson P (1999) Natural poly-histidine affinity tag for purification of recombinant proteins on cobalt(II)-carboxymethylaspartate crosslinked agarose. Journal of Chromatography A, 864*:*247–256*.*

12. Chaga G, Hopp J, Nelson P (1999) Immobilized metal ion affinity chromatography on Co2+ carboxymethylaspartate–agarose Superflow, as demonstrated by one step purification of

lactate dehydrogenase from chicken breast muscle. Biotechnology and Applied Biochemistry, 29:19-24.

13. Clay K, Schardl C (2002) Evolutionary origins and ecological consequences of endophyte symbiosis with grasses. The American Naturalist 160:99–127.

14. Dodge AG, Wackett LP (2005) Metabolism of bismuth subsalicylate and intracellular accumulation of bismuth by *Fusarium* sp. strain BI. Applied and Environmental Microbiology, 71: 876-882.

15. Durrant WE, Dong X (2004) Systemic acquired resistance. Annual Reviews of Phytopathology, 42:185-209.

16. Esposito D, Chatterjee DK (2006) Enhancement of soluble protein expression through the use of fusion tags. Current Opinion in Biotechnology, 17: 353-358.

17. Eppink MH, van Berkel WJ, Schreuder HA (1997) Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. Protein Science, 6:2454-2458.

18. Eppink MH, Schreuder HA, van Berkel WJ (1998). Interdomain binding of NADPH in p-hydroxybenzoate hydroxylase as suggested by kinetic, crystallographic and modeling studies of histidine 162 and arginine 269 variants. Journal of Biological Chemistry, 273: 21031-21039.

19. Ferrer M, Chernikova TN, Yakimov MM, Golyshin PN, Timmis KN (2003) Chaperonins govern growth of *Escherichia coli* at low temperatures. Nature Biotechnology, 21:1266-1267.

20. Finzi A, Cloutier J, Cohen ÉA (2003) Two-step purification of His-tagged Nef protein in native condition using heparin and immobilized metal ion affinity chromatographies. Journal of Virological Methods, 111:69-73.

21. Gaffney T, Friedrich L, Vernooij B, Negrotto D, Nye G, Uknes S, Ward E, Kessmann H, Ryals J (1993) Requirement of salicylic acid for the induction of systemic acquired resistance. Science, 261:754-756.

22. Galloway CA, Sowden MP, Smith HC (2003) Increasing the yield of soluble recombinant protein expressed in *E. coli* by induction during late log phase. Biotechniques, 34:524-6.

23. Graminha MA, Rocha EM, Prade RA, Martinez-Rossi NM (2004) Terbinafine resistance mediated by salicylate 1-monooxygenase in *Aspergillus nidulans*. Antimicrobial Agents and Chemotherapy, 48:3530-3535.

24. Goujon M, McWilliam H, Li W, Valentin F, Squizzato S, Paern J, Lopez R (2010) A new bioinformatics analysis tools framework at EMBL–EBI. Nucleic Acids Research, 38:695-699.

25. Hartinger D, Heinl S, Schwartz HE, Grabherr R, Schatzmayr G, Haltrich D, Moll WD (2010) Enhancement of solubility in *Escherichia coli* and purification of an aminotransferase from *Sphingopyxis* sp. MTA 144 for deamination of hydrolyzed fumonisin B(1). Microbial Cell Factories, 9:62-62.

26. Heck S, Grau T, Buchala A, Métraux JP, Nawrath C (2003) Genetic evidence that expression of NahG modifies defence pathways independent of salicylic acid biosynthesis in the *Arabidopsis–Pseudomonas syringae* pv. tomato interaction. The Plant Journal, 36: 342-352.

27. Hottenrott S, Schumann T, Plückthun A, Fischer G, Rahfeld JU (1997) The *Escherichia coli* SlyD is a metal ion-regulated peptidyl-prolyl cis/trans-isomerase. Journal of Biological Chemistry, 25:15697-15701.

28. Iwasaki Y, Gunji H, Kino K, Hattori T, Ishii Y, Kirimura K (2010) Novel metabolic pathway for salicylate biodegradation via phenol in yeast *Trichosporon moniliiforme*. Biodegradation, 21:557-564.

29. Keren H, Lev-Maor G, Ast G (2010) Alternative splicing and evolution: diversification, exon definition and function. Nature Reviews Genetics, 11:345-355.

30. Kohler HP, Schmid A, van der Maarel M (1993) Metabolism of 2,2'-dihydroxybiphenyl by Pseudomonas sp. strain HBP1: production and consumption of 2,2',3-trihydroxybiphenyl. Journal of Bacteriology, 175:1621-8.

31. Kopetzki E, Schumacher G, Buckel P (1989) Control of formation of active soluble or inactive insoluble baker's yeast α-glucosidase PI in *Escherichia coli* by induction and growth conditions. Molecular and General Genetics, 216: 149-155.

32. Kornblihtt AR, Schor IE, Alló M, Dujardin G, Petrillo E, Muñoz MJ (2013) Alternative splicing: a pivotal step between eukaryotic transcription and translation. Nature Reviews Molecular Cell Biology, 14:153-165.

33. Kunkel BN, Brooks DM (2002) Cross talk between signaling pathways in pathogen defense. Current Opinion in Plant Biology, 5:325-331.

34. Lichty JJ, Malecki JL, Agnew HD, Michelson-Horowitz DJ, Tan S (2005) Comparison of affinity tags for protein purification. Protein Expression and Purification, 41: 98-105.

35. Lilie H, Schwarz E, Rudolph R (1998) Advances in refolding of proteins produced in *E. coli*. Current Opinion in Biotechnology, 9:497-501.

36. Lin FH, Sun T, Fletcher GL, Davies PL (2012) Thermolabile antifreeze protein produced in *Escherichia coli* for structural analysis. Protein Expression and Purification, 82:75-82.

37. Maratea D, Young K, Young R *(*1985*)* Deletion and fusion analysis of the phage phi X174 lysis gene E. Gene 40*:*39–46.

38. Marshall AN, Montealegre MC, Jiménez-López C, Lorenz MC, van Hoof A (2013) Alternative splicing and subfunctionalization generates functional diversity in fungal proteomes. PLoS Genetics, 9:e1003376.

39. Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: towards a cellular code. Nature Reviews Molecular Cell Biology, 6:386-398.

40. Métraux JP, Signer H, Ryals J, Ward E, Wyss-Benz M, Gaudin J, Raschdorf K, Schmid E, Blum W, Inverardi B (1990) Increase in salicylic acid at the onset of systemic acquired

resistance in cucumber. Science, 250:1004–6.

41. McGuire AM, Pearson MD, Neafsey DE, Galagan JE (2008) Cross-kingdom patterns of alternative splicing and splice recognition. Genome Biology, 9:R50.

42. McMurry JL, Macnab RM (2004) Cobalt-based protein purification resin does not bind *E. coli* SlyD, a common contaminant in Ni-NTA IMAC. Advantage Business Media, Rockaway, NJ.

43. Mitterauer T, Nanoff C, Ahorn H, Freissmuth M, Hohenegger M (1999) Metal-dependent nucleotide binding to the *Escherichia coli* rotamase SlyD. Biochemical Journal, 342:33-39.

44. Moghaddam A, Reinton N (2008) Use of polyethylene glycol for drying polyacrylamide gels to avoid cracking. SciPhu Publishing. Retrieved from: http://sciphu.com/2008/03/use-of-polyethylene-glycol-for-drying.html

45. Morse AM, Tschaplinski TJ, Dervinis C, Pijut PM, Schmelz EA, Day W, Davis JM (2007) Salicylate and catechol levels are maintained in *nahG* transgenic poplar. Phytochemistry, 68: 2043-2052.

46. Mukherjee S, Shukla A, Guptasarma P (2003) Single‐step purification of a protein‐folding catalyst, the SlyD peptidyl prolyl isomerase (PPI), from cytoplasmic extracts of *Escherichia coli*. Biotechnology and Applied Biochemistry, 2:183-186.

47. Pacheco B, Crombet L, Loppnau P, Cossar D (2012) A screening strategy for heterologous protein expression in *Escherichia coli* with the highest return of investment. Protein Expression and Purification, 81:33-41.

48. Parsy CB, Chapman CJ, Barnes AC, Robertson JF, Murray A (2007) Two-step method to isolate target recombinant protein from co-purified bacterial contaminant SlyD after immobilised metal affinity chromatography. Journal of Chromatography B, 853:314-319.

49. Pieterse CMJ, Van der Does D, Zamioudis C, Leon-Reyes A, Van Wees SCM (2012) Hormonal modulation of plant immunity. Annual Review of Cell and Developmental Biology, 28:489-521.

50. Phan TC, Nowak KJ, Anthony Akkari P, Zheng MH, Xu J (2003) Expression of caltrin in the baculovirus system and its purification in high yield and purity by cobalt (II) affinity chromatography. Protein Expression and Purification, 29:284-290.

51. Purbey PK, Jayakumar PC, Patole MS, Galande S (2006) pC6-2/caspase-6 system to purify glutathione-S-transferase-free recombinant fusion proteins expressed in *Escherichia coli*. Nature Protocols, 1:1820-1827.

52. Rabe F, Ajami-Rashidi Z, Doehlemann G, Kahmann R, Djamei A (2013) Degradation of the plant defence hormone salicylic acid by the biotrophic fungus *Ustilago maydis*. Molecular Microbiology, 89:179-188.

53. Raskin I (1992) Salicylate, a new plant hormone. Plant Physiology, 99:799.

54. Robichon C, Luo J, Causey TB, Benner JS, Samuelson JC (2011) Engineering *Escherichia coli* BL21 (DE3) derivative strains to minimize *E. coli protein* contamination after purification by immobilized metal affinity chromatography. Applied and Environmental Microbiology, 77:4634-4646.

55. Roof WD, Horne SM, Young KD, Young R (1994) slyD, a host gene required for phi X174 lysis, is related to the FK506-binding protein family of peptidyl-prolyl cis-trans-isomerases. Journal of Biological Chemistry, 4:2902-2910.

56. Roof WD, Fang HQ, Young KD, Sun J, Young R (1997). Mutational analysis of slyD, an *Escherichia coli* gene encoding a protein of the FKBP immunophilin family. Molecular Microbiology 6:1031-1046.

57. Schardl CL, Young CA, Hesse U, Amyotte SG, Andreeva K, Calie PJ, Fleetwood DJ, Haws DC, Moore N, Oeser B, Panaccione DG, Schweri KK, Voisey CR, Farman ML, Jaromczyk JW, Roe BA, O'Sullivan DM, Scott B, Tudzynski P, An Z, Arnaoudova EG, Bullock CT, Charlton ND, Chen L, Cox M, Dinkins RD, Florea S, Glenn AE, Gordon A, Güldener U, Harris DR, Hollin W, Jaromczyk J, Johnson RD, Khan AK, Leistner E, Leuchtmann A, Li C, Liu J, Liu J, Liu M, Mace W, Machado C, Nagabhyru P, Pan J, Schmid J, Sugawara K, Steiner U, Takach J, Tanaka E, Webb JS, Wilson EV, Wiseman JL, Yoshida R, Zeng Z. (2013) Plant-symbiotic fungi as chemical engineers: multi-genome analysis of the *Clavicipitaceae* reveals dynamics of alkaloid loci. PLoS Genetics, 9(2):e1003323.

58. Scholz C, Eckert B, Hagn F, Schaarschmidt P, Balbach J, Schmid FX (2006) SlyD proteins from different species exhibit high prolyl isomerase and chaperone activities. Biochemistry, 1:20-33.

59. Seo JS, Keum YS, Li QX (2009) Bacterial degradation of aromatic compounds. International Journal of Environmental Research and Public Health, 6:278-309.

60. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, Thompson JD, Higgins, DG (2011) Fast, scalable generation of high quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7 doi: 10.1038/msb.2011.75.

61. Strandberg L, Enfors SO (1991) Factors influencing inclusion body formation in the production of a fused protein in *Escherichia coli*. Applied and Environmental Microbiology, 57:1669-1674.

62. Suemori A (2013) Conserved and non-conserved residues and their role in the structure and function of p-hydroxybenzoate hydroxylase. Protein Engineering Design and Selection, 26:479-488.

63. Suzuki K, Asao E, Nakamura Y, Nakamura M, Ohnishi K, Fukudat S (2000) Overexpression of salicylate hydroxylase and the crucial role of Lys163 as its NADH binding site. Journal of Biochemistry, 128:293-299.

64. Szu IS-Y, Dagley S (1984) Properties of salicylate hydroxylase and hydroxyquinol 1,2-dioxygenase purified from *Trichosporon cutaneum*. Journal of Bacteriology, 159:353-359.

65. Van Wees S, Glazebrook J (2003) Loss of non‐host resistance of *Arabidopsis* NahG to *Pseudomonas syringae* pv. *phaseolicola* is due to degradation products of salicylic acid. The Plant Journal, 33: 733-742.

66. Wülfing C, Lombardero J, Plückthun A (1994) An *Escherichia coli* protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif. Journal of Biological Chemistry, 4:2895-2901.

67. Yamamoto S, Katagiri M, Maeno H, Hayaishi O (1965) Salicylate hydroxylase, a monooxygenase requiring flavin adenine dinucleotide I. Purification and general properties. Journal of Biological Chemistry, 240:3408–3413.

68. You IS, Murray RI, Jollie D, Gunsalus IC (1990) Purification and characterization of salicylate hydroxylase from *Pseudomonas putida* PpG7. Biochemical and Biophysical Research Communications, 169:1049-1054.

69. Zhao H, Chen D, Li Y, Cai B (2005) Overexpression, purification and characterization of a new salicylate hydroxylase from naphthalene-degrading *Pseudomonas* sp. strain ND6. Microbiological Research, 160:307-313.

Chapter 5

Summary

Chapter 5

Summary

The overall goal of this thesis was to use SOLiD-SAGE transcriptome analysis to identify candidate genes that may be important in the endophyte-grass symbiosis. As discussed in previous chapters, the fungal endophyte-turfgrass mutualism is highly beneficial to the host plants because of the positive effects the endophyte confers to host fitness. The *E. festucae- F. rubra* (strong creeping red fescue) interaction is of special importance in the work described in this thesis.

Fine fescues-infected with *E. festucae* have been shown to be resistant to insect herbivory, and two important fungal diseases. The diseases are dollar spot, which is the most costly infection to treat in the management of turf, and red thread pathogenesis. To date, no other *Epichloë* spp. infection of cool-season turfgrass has been observed to confer the unique endophyte-mediated fungal disease resistance. Therefore protection against fungal diseases that is observed in the *E. festucae*-fine fescue mutualism provided a compelling rationale to examine the interaction to identify candidate molecules involved in disease resistance.

The first project described in this thesis revealed the differential expression of over 200 host transcripts that moderately impact many plant physiological processes as a result of *E. festucae* endophyte infection. The study also showed that the abundance of a transcript encoding an antifungal protein that may play a role in the fungal disease resistance observed in *E. festucae*-infected red fescue. There were many endophyte transcripts identified that encode secreted small fungal proteins that possibly play a role in the regulation of the fungal-plant symbiosis, and contribute to the persistence of the interaction.

Several endophyte transcripts identified in the transcriptome profiling were candidates for

encoding proteins involved in the evasion and/or suppression of plant host defense. One such transcript encoding fungal salicylate hydroxylase was examined to test this hypothesis. An endophyte insect toxin gene, acquired by horizontal gene transfer from a bacterium, was also characterized. The toxin may be a component, in addition to the fungal alkaloids, in the insect resistance observed in endophyte-infected grasses. Therefore, the studies discussed here have contributed to achieve the overall objective of this thesis and increase the current knowledge of *Epichloë*-turfgrass mutualistic symbiotic interactions.